

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 March 2002 (28.03.2002)

PCT

(10) International Publication Number  
**WO 02/24895 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 15/00**

(21) International Application Number: PCT/US01/29834

(22) International Filing Date:  
21 September 2001 (21.09.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/234,903 22 September 2000 (22.09.2000) US  
60/244,505 30 October 2000 (30.10.2000) US  
60/254,402 8 December 2000 (08.12.2000) US

(71) Applicant (*for all designated States except US*): **INCYTE GENOMICS, INC.** [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **NGUYEN, Dannel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). **HALAFIA, April, J., A.** [US/US]; 2227 Calle de Primavera, Santa Clara, Ca 95054 (US). **WALIA, Narinder, K.** [US/US]; 890 Davis Street #205, San Leandro, CA 94577 (US). **YAO, Monique, G.** [US/US]; 1189 Woodgate Drive, Carmel, IN 46033 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). **RAMKUMAR, Jayalaxmi** [IN/US]; 34359 Maybird Circle, Fremont, CA 94555 (US). **THANGAVELU, Kavitha** [IN/US]; 1950 Montecito Avenue #23, Mountain View, CA 94304 (US). **LU, Yan** [CN/US]; 3885 Corrina Way, Palo Alto,

CA 94303 (US). **LEE, Sally** [CN/US]; 825 East Evelyn, #425, Sunnyvale, CA 94086 (US). **BAUGHN, Mariah, R.** [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). **TANG, Y., Tom** [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **AZIMZAI, Yalda** [US/US]; 5518 Boulder Canyon Drive, Castro Valley, CA 94552 (US). **KALAFUS, Daniel, P.** [US/US]; 551 25th Avenue, San Francisco, CA 94121 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US).

(74) Agents: **HAMLET-COX, Diana** et al.; Incyte Genomics, Inc, 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *without international search report and to be republished upon receipt of that report*

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: TRANSCRIPTION FACTORS AND ZINC FINGER PROTEINS

(57) Abstract: The invention provides human transcription factors and zinc finger proteins (TFZN) and polynucleotides which identify and encode TFZN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TFZN.



WO 02/24895 A2

## TRANSCRIPTION FACTORS AND ZINC FINGER PROTEINS

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transcription factors and  
5 zinc finger proteins and to the use of these sequences in the diagnosis, treatment, and prevention of  
cell proliferative disorders, including cancer, and developmental, autoimmune/inflammatory and  
neurological disorders, and in the assessment of the effects of exogenous compounds on the  
expression of nucleic acid and amino acid sequences of transcription factors and zinc finger proteins.

### 10 BACKGROUND OF THE INVENTION

Multicellular organisms are comprised of diverse cell types that differ dramatically both in  
structure and function. The identity of a cell is determined by its characteristic pattern of gene  
expression, and different cell types express overlapping but distinctive sets of genes throughout  
development. Spatial and temporal regulation of gene expression is critical for the control of cell  
15 proliferation, cell differentiation, apoptosis, and other processes that contribute to organismal  
development. Furthermore, gene expression is regulated in response to extracellular signals that  
mediate cell-cell communication and coordinate the activities of different cell types. Appropriate gene  
regulation also ensures that cells function efficiently by expressing only those genes whose functions  
are required at a given time.

20 Transcriptional regulatory proteins are essential for the control of gene expression. Some of  
these proteins function as transcription factors that initiate, activate, repress, or terminate gene  
transcription. Transcription factors generally bind to the promoter, enhancer, and upstream regulatory  
regions of a gene in a sequence-specific manner, although some factors bind regulatory elements  
within or downstream of a gene's coding region. Transcription factors may bind to a specific region of  
25 DNA singly or as a complex with other accessory factors. (Reviewed in Lewin, B. (1990) Genes IV,  
Oxford University Press, New York, NY, and Cell Press, Cambridge, MA, pp. 554-570.)

The double helix structure and repeated sequences of DNA create topological and chemical  
features which can be recognized by transcription factors. These features are hydrogen bond donor  
and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches  
30 of sequence which induce distinct bends in the helix. Typically, transcription factors recognize specific  
DNA sequence motifs of about 20 nucleotides in length. Multiple, adjacent transcription factor-binding  
motifs may be required for gene regulation.

Many transcription factors incorporate DNA-binding structural motifs which comprise either  
 $\alpha$  helices or  $\beta$  sheets that bind to the major groove of DNA. Four well-characterized structural motifs

are helix-turn-helix, zinc finger, leucine zipper, and helix-loop-helix. Proteins containing these motifs may act alone as monomers, or they may form homo- or heterodimers that interact with DNA.

The helix-turn-helix motif consists of two  $\alpha$  helices connected at a fixed angle by a short chain of amino acids. One of the helices binds to the major groove. Helix-turn-helix motifs are exemplified by the homeobox motif which is present in homeodomain proteins. These proteins are critical for specifying the anterior-posterior body axis during development and are conserved throughout the animal kingdom. The Antennapedia and Ultrabithorax proteins of Drosophila melanogaster are prototypical homeodomain proteins. (Pabo, C.O. and R.T. Sauer (1992) Ann. Rev. Biochem. 61:1053-1095.)

The zinc finger motif, which binds zinc ions, generally contains tandem repeats of about 30 amino acids consisting of periodically spaced cysteine and histidine residues. Examples of this sequence pattern, designated C2H2 and C3HC4 ("RING" finger), zinc fingers, and the PHD domain (Lewin, supra ; Aasland, R. et al. (1995) Trends Biochem. Sci 20:56 - 59). Zinc finger proteins each contain an  $\alpha$  helix and an antiparallel  $\beta$  sheet whose proximity and conformation are maintained by the zinc ion. Contact with DNA is made by the arginine preceding the  $\alpha$  helix and by the second, third, and sixth residues of the  $\alpha$  helix. The zinc finger motif may be repeated in a tandem array within a protein, such that the  $\alpha$  helix of each zinc finger in the protein makes contact with the major groove of the DNA double helix. This repeated contact between the protein and the DNA produces a strong and specific DNA-protein interaction. The strength and specificity of the interaction can be regulated by the number of zinc finger motifs within the protein. Variants of the zinc finger motif include poorly defined cysteine-rich motifs which bind zinc or other metal ions. These motifs may not contain histidine residues and are generally non-repetitive. Zinc-finger transcription factors are often accompanied by modular sequence motifs such as the Kruppel-associated box (KRAB) and the SCAN domain. The KRAB domain mediates transcriptional repression while the SCAN domain mediates selective protein dimerization. The hypoalphalipoproteinemia susceptibility gene ZNF202 encodes a SCAN box and a KRAB domain followed by eight C2H2 zinc-finger motifs (Honer, C. et al. (2001) Biochim. Biophys. Acta 1517:441-448).

The leucine zipper motif comprises a stretch of amino acids rich in leucine which can form an amphipathic  $\alpha$  helix. This structure provides the basis for dimerization of two leucine zipper proteins. The region adjacent to the leucine zipper is usually basic, and upon protein dimerization, is optimally positioned for binding to the major groove. Proteins containing such motifs are generally referred to as bZIP transcription factors. The leucine zipper motif is found in the proto-oncogenes Fos and Jun, which comprise the heterodimeric transcription factor AP1, involved in cell growth and the determination of cell lineage (Papavassiliou, A. G. (1995) N. Engl. J. Med. 332:45-47).

The helix-loop-helix motif (HLH) consists of a short  $\alpha$  helix connected by a loop to a longer  $\alpha$  helix. The loop is flexible and allows the two helices to fold back against each other and to bind to DNA. The oncogene Myc, a transcription factor that activates genes required for cellular proliferation, contains a prototypical HLH motif.

5        The NF-kappa-B/Rel signature defines a family of eukaryotic transcription factors involved in oncogenesis, embryonic development, differentiation and immune response. Most transcription factors containing the Rel homology domain (RHD) bind as dimers to a consensus DNA sequence motif termed kappa-B. Members of the Rel family share a highly conserved 300 amino acid domain termed the Rel homology domain. The characteristic Rel C-terminal domain is involved in gene activation and  
10        cytoplasmic anchoring functions. Proteins known to contain the RHD domain include vertebrate nuclear factor NF-kappa-B, which is a heterodimer of a DNA-binding subunit and the transcription factor p65, mammalian transcription factor RelB, and vertebrate proto-oncogene c-rel, a protein associated with differentiation and lymphopoiesis (Kabrunk, N., and Enrietto, P.J. (1994) *Semin. Cancer Biol.* 5:103-112).

15        A DNA binding motif termed ARID (AT-rich interactive domain) distinguishes an evolutionarily conserved family of proteins. The approximately 100-residue ARID sequence is present in a series of proteins strongly implicated in the regulation of cell growth, development, and tissue-specific gene expression. ARID proteins include Bright (a regulator of B-cell-specific gene expression), dead ringer (involved in development), and MRF-2 (which represses expression from the  
20        cytomegalovirus enhancer) (Dallas, P.B. et al. (2000) *Mol. Cell Biol.* 20:3137-3146).

The ELM2 (Egl-27 and MTA1 homology 2) domain is found in metastasis-associated protein MTA1 and protein ER1. The Caenorhabditis elegans gene egl-27 is required for embryonic patterning MTA1, a human gene with elevated expression in metastatic carcinomas, is a component of a protein complex with histone deacetylase and nucleosome remodelling activities (Solari, F. et al. (1999)  
25        *Development* 126:2483-2494). The ELM2 domain is usually found to the N terminus of a myb-like DNA binding domain. ELM2 is also found associated with an ARID DNA

Most transcription factors contain characteristic DNA binding motifs, and variations on the above motifs and new motifs have been and are currently being characterized. (Faisst, S. and S. Meyer (1992) *Nucl. Acids Res.* 20:3-26.) These include the forkhead motif, found in transcription  
30        factors involved in development and oncogenesis (Hacker, U. et al. (1995) *EMBO J.* 14:5306-5317).

Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes. (Cleary, M.L. (1992) *Cancer Surv.* 15:89-104.) Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one



gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in inappropriate gene transcription, potentially contributing to malignancy.

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient regulation of gene expression may result in considerable tissue or organ damage. This damage is well documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections. (Isselbacher et al. Harrison's Principles of Internal Medicine, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996.)

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development could result in developmental disorders.

#### Chromatin Associated Proteins

In the nucleus, DNA is packaged into chromatin, the compact organization of which limits the accessibility of DNA to transcription factors and plays a key role in gene regulation (Lewin, supra, pp. 409-410). The compact structure of chromatin is determined and influenced by chromatin-associated proteins such as the histones, the high mobility group (HMG) proteins, helicases, and the chromodomain proteins. There are five classes of histones, H1, H2A, H2B, H3, and H4, all of which are highly basic, low molecular weight proteins. The fundamental unit of chromatin, the nucleosome, consists of 200 base pairs of DNA associated with two copies each of H2A, H2B, H3, and H4. H1 links adjacent nucleosomes. HMG proteins are low molecular weight, non-histone proteins that may play a role in unwinding DNA and stabilizing single-stranded DNA. Helicases, which are DNA-dependent ATPases, unwind DNA, allowing access for transcription factors. Chromodomain proteins play a key role in the formation of highly compacted heterochromatin, which is transcriptionally silent.

#### Zinc Finger Proteins

A zinc finger is a cysteine-rich, compactly folded protein motif in which specifically placed cysteines, and in some cases histidines, coordinate  $Zn^{+2}$ . Several types of zinc finger motifs have been identified. Though originally identified in DNA-binding proteins as regions that interact directly with DNA, zinc fingers occur in a variety of proteins that do not bind DNA (Lodish, H. et al. (1995) Molecular Cell Biology, Scientific American Books, New York, NY, pp. 447-451). For example, Galcheva-Gargova, Z. et al. ((1996) *Science* 272:1797-1802) have identified zinc finger proteins that interact with various cytokine receptors.

The C2H2-type zinc finger signature motif contains a 28 amino acid sequence, including 2 conserved Cys and 2 conserved His residues in a C-2-C-12-H-3-H type motif. The motif generally occurs in multiple tandem repeats. Zinc finger proteins each contain an  $\alpha$  helix and an antiparallel  $\beta$  sheet, whose proximity and conformation are maintained by the zinc ion. Examples of zinc finger proteins include the C2H2-type and DHHC, the C4-type, the C3HC4-type zinc fingers, and the GATA domain (Lewin, *supra*; Aasland, R. *et al.* (1995) Trends Biochem. Sci. 20:56-59). A cysteine-rich domain including the motif Asp-His-His-Cys (DHHC-CRD) has been identified as a distinct subgroup of zinc finger proteins. The DHHC-CRD region has been implicated in growth and development. One DHHC-CRD mutant shows defective function of Ras, a small membrane-associated GTP-binding protein that regulates cell growth and differentiation, while other DHHC-CRD proteins probably function in pathways not involving Ras (Bartels, D.J. *et al.* (1999) Mol. Cell Biol. 19:6775-6787).

The SCAN domain is a highly conserved, leucine-rich motif of approximately 60 amino acids found at the amino-terminal end of zinc finger transcription factors. SCAN domains are most often linked to C2H2 zinc finger motifs through their carboxyl-terminal end. Biochemical binding studies have established the SCAN domain as a selective hetero- and homotypic oligomerization domain. SCAN domain-mediated protein complexes may function to modulate the biological function of transcription factors (Schumacher, C. *et al.*, (2000) J. Biol. Chem. 275:17173-17179).

The KRAB (Kruppel-associated box) domain is a conserved amino acid sequence spanning approximately 75 amino acids and is found in almost one-third of the 300 to 700 genes encoding C2H2 zinc fingers. The KRAB domain is generally encoded by two exons. The KRAB-A region or box is encoded by one exon and the KRAB-B region or box is encoded by a second exon. The function of the KRAB domain is the repression of transcription. Transcription repression is accomplished by recruitment of either the KRAB-associated protein-1, a transcriptional corepressor, or the KRAB-A interacting protein. Proteins containing the KRAB domain are likely to play a regulatory role during development (Williams, A.J. *et al.*, (1999) Mol. Cell Biol. 19:8526-8535). KRAB (Kruppel-associated box) is an evolutionarily conserved protein domain found N-terminally with respect to the finger repeats. A subgroup of highly related human KRAB zinc finger proteins detectable in all human tissues is highly expressed in human T lymphoid cells (Bellefroid, E.J. *et al.* (1993) EMBO J. 12:1363-1374). The ZNF85 KRAB zinc finger gene, a member of the human ZNF91 family, is highly expressed in normal adult testis, in seminomas, and in the NT2/D1 teratocarcinoma cell line (Poncelet, D.A. *et al.* (1998) DNA Cell Biol. 17:931-943).

The C4 motif is found in hormone-regulated proteins. The C4 motif generally includes only 2 repeats. A number of eukaryotic and viral proteins contain a conserved cysteine-rich domain of 40

to 60 residues (called C3HC4 zinc-finger or RING finger) that binds two atoms of zinc, and is probably involved in mediating protein-protein interactions. The 3D "cross-brace" structure of the zinc ligation system is unique to the RING domain. The spacing of the cysteines in such a domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C.

5       The PHD finger is a C4HC3 zinc-finger-like motif found in nuclear proteins thought to be involved in chromatin-mediated transcriptional regulation. Transcriptional regulatory proteins control gene expression by activating or repressing gene transcription. Transcription factors generally bind to regulatory regions of a gene in a sequence-specific manner usually in the promoter or enhancer region upstream of the coding sequence. Transcription factors recognize topological and chemical features  
10 such as hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular repeated stretches of sequence which induce distinct bends in the helix. Multiple adjacent transcription factor-binding motifs may be required for gene regulation. (Reviewed in Lewin, B. (1990) Genes IV, Oxford University Press, New York, NY, pp. 554-570.)

GATA-type transcription factors contain one or two zinc finger domains which bind  
15 specifically to a region of DNA that contains the consecutive nucleotide sequence GATA. The zinc finger domain consensus sequence is C-X(2)-C-X(4,8)-W-X(9,10)-C-X(2)-C, wherein X is any amino acid, and the numbers in the parentheses indicate the range in the number of amino acids within that region. NMR studies indicate that the zinc finger comprises two irregular anti-parallel  $\beta$  sheets and an  $\alpha$  helix, followed by a long loop to the C-terminal end of the finger (Ominchinski, J.G. (1993) *Science*  
20 261:438-446). The helix and the loop connecting the two  $\beta$ -sheets contact the major groove of the DNA, while the C-terminal part, which determines the specificity of binding, wraps around into the minor groove.

The LIM motif consists of about 60 amino-acid residues and contains seven conserved cysteine residues and a histidine within a consensus sequence (Schmeichel, K.L. and Beckerle, M.C.  
25 (1994) *Cell* 79:211-219). The LIM family includes transcription factors and cytoskeletal proteins which may be involved in development, differentiation, and cell growth. One example is actin-binding LIM protein, which may play roles in regulation of the cytoskeleton and cellular morphogenesis (Roof, D.J. et al. (1997) *J. Cell Biol.* 138:575-588). The N-terminal domain of actin-binding LIM protein has four double zinc finger motifs with the LIM consensus sequence. The C-terminal domain of actin-  
30 binding LIM protein shows sequence similarity to known actin-binding proteins such as dematin and villin. Actin-binding LIM protein binds to F-actin through its dematin-like C-terminal domain. The LIM domain may mediate protein-protein interactions with other LIM-binding proteins.

Myeloid cell development is controlled by tissue-specific transcription factors. Myeloid zinc finger proteins (MZF) include MZF-1 and MZF-2. MZF-1 functions in regulation of the development

of neutrophilic granulocytes. A murine homolog MZF-2 is expressed in myeloid cells, particularly in the cells committed to the neutrophilic lineage, and down-regulated by G-CSF and appears to have a unique function in neutrophil development (Murai, K. et al. (1997) *Genes Cells* 2:581-591).

#### Diseases and disorders related to gene regulation

5 Many neoplastic disorders in humans can be attributed to inappropriate gene expression. Malignant cell growth may result from either excessive expression of tumor promoting genes or insufficient expression of tumor suppressor genes (Cleary, M.L. (1992) *Cancer Surv.* 15:89-104). Chromosomal translocations may also produce chimeric loci which fuse the coding sequence of one gene with the regulatory regions of a second unrelated gene. Such an arrangement likely results in  
 10 inappropriate gene transcription, potentially contributing to malignancy. One clinically relevant zinc-finger protein is WT1, a tumor-suppressor protein that is inactivated in children with Wilm's tumor. The oncogene bcl-6, which plays an important role in large-cell lymphoma, is also a zinc-finger protein (Papavassiliou, *supra*). In Burkitt's lymphoma, for example, the transcription factor Myc is translocated to the immunoglobulin heavy chain locus, greatly enhancing Myc expression and resulting  
 15 in rapid cell growth leading to leukemia (Latchman, D. S. (1996) *N. Engl. J. Med.* 334:28-33).

In addition, the immune system responds to infection or trauma by activating a cascade of events that coordinate the progressive selection, amplification, and mobilization of cellular defense mechanisms. A complex and balanced program of gene activation and repression is involved in this process. However, hyperactivity of the immune system as a result of improper or insufficient  
 20 regulation of gene expression may result in considerable tissue or organ damage. This damage is well documented in immunological responses associated with arthritis, allergens, heart attack, stroke, and infections (Isselbacher et al. *Harrison's Principles of Internal Medicine*, 13/e, McGraw Hill, Inc. and Teton Data Systems Software, 1996). The causative gene for autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) was recently isolated and found to encode a protein  
 25 with two PHD-type zinc finger motifs (Bjorses, P. et al. (1998) *Hum. Mol. Genet.* 7:1547-1553).

Furthermore, the generation of multicellular organisms is based upon the induction and coordination of cell differentiation at the appropriate stages of development. Central to this process is differential gene expression, which confers the distinct identities of cells and tissues throughout the body. Failure to regulate gene expression during development can result in developmental disorders.  
 30 Human developmental disorders caused by mutations in zinc finger-type transcriptional regulators include: urogenital developmental abnormalities associated with WT1; Greig cephalopolysyndactyly, Pallister-Hall syndrome, and postaxial polydactyly type A (GLI3); and Townes-Brocks syndrome, characterized by anal, renal, limb, and ear abnormalities (SALL1) (Engelkamp, D. and van Heyningen, V. (1996) *Curr. Opin. Genet. Dev.* 6:334-342; Kohlhase, J. et al. (1999) *Am. J. Hum. Genet.* 64:435-

445).

Impaired transcriptional regulation may lead to Alzheimer's disease, a progressive neurodegenerative disorder that is characterized by the formation of senile plaques and neurofibrillary tangles containing amyloid beta peptide. These plaques are found in limbic and association cortices of the brain, including hippocampus, temporal cortices, cingulate cortex, amygdala, nucleus basalis and locus caeruleus. Early in Alzheimer's pathology, physiological changes are visible in the cingulate cortex (Minoshima, S. et al. (1997) *Annals of Neurology* 42:85-94). In subjects with advanced Alzheimer's disease, accumulating plaques damage the neuronal architecture in limbic areas and eventually cripple the memory process.

The discovery of new transcription factors and zinc finger proteins, and the polynucleotides encoding them, satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cell proliferative disorders, including cancer, and developmental, autoimmune/inflammatory and neurological disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transcription factors and zinc finger proteins.

### SUMMARY OF THE INVENTION

The invention features purified polypeptides, transcription factors and zinc finger proteins, referred to collectively as "TFZN" and individually as "TFZN-1," "TFZN-2," "TFZN-3," "TFZN-4," "TFZN-5," "TFZN-6," "TFZN-7," and "TFZN-8." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an

amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-8. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:9-16.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter  
5 sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group  
10 consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group  
15 consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid  
20 sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a  
25 polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic  
30 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least

90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

5            Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c)  
10 a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said  
15 probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

            The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a  
20 polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said  
25 target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

            The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected  
30 from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and

a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional TFZN, comprising administering to a patient in need of such treatment the composition.

5       The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having  
10 an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a  
15 pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TFZN, comprising administering to a patient in need of such treatment the composition.

      Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide  
20 comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group  
25 consisting of SEQ ID NO:1-8. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TFZN, comprising  
30 administering to a patient in need of such treatment the composition.

      The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected



from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one  
5 test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, b) a polypeptide comprising a  
10 naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8. The method comprises a) combining the polypeptide with at least one  
15 test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

20 The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

25 The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a  
30 polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological

sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

### DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing

particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to “a host cell” includes a plurality of such host cells, and a reference to “an antibody” is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs.

Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

“TFZN” refers to the amino acid sequences of substantially purified TFZN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term “agonist” refers to a molecule which intensifies or mimics the biological activity of TFZN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TFZN either by directly interacting with TFZN or by acting on components of the biological pathway in which TFZN participates.

An “allelic variant” is an alternative form of the gene encoding TFZN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

“Altered” nucleic acid sequences encoding TFZN include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TFZN or a polypeptide with at least one functional characteristic of TFZN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of

the polynucleotide encoding TFZN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TFZN. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TFZN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TFZN is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TFZN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TFZN either by directly interacting with TFZN or by acting on components of the biological pathway in which TFZN participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TFZN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an in vitro evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries.

Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH<sub>2</sub>), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker. (See, e.g., Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13.)

The term "intramer" refers to an aptamer which is expressed in vivo. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the

designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TFZN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TFZN or fragments of TFZN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His

	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
5	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
10	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

---

15       Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

20       A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

      The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

      A “detectable label” refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

30       “Differential expression” refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

      “Exon shuffling” refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

      A “fragment” is a unique portion of TFZN or the polynucleotide encoding TFZN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up

to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:9-16 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:9-16, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:9-16 is useful, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:9-16 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:9-16 and the region of SEQ ID NO:9-16 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-8 is encoded by a fragment of SEQ ID NO:9-16. A fragment of SEQ ID NO:1-8 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-8. For example, a fragment of SEQ ID NO:1-8 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-8. The precise length of a fragment of SEQ ID NO:1-8 and the region of SEQ ID NO:1-8 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A “full length” polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, interchangeably, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.



Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Reward for match: 1*

*Penalty for mismatch: -2*

*Open Gap: 5 and Extension Gap: 2 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 11*

*Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous

nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

5 Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

10 The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

15 Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by  
20 CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

25       *Matrix: BLOSUM62*  
           *Open Gap: 11 and Extension Gap: 1 penalties*  
           *Gap x drop-off: 50*  
           *Expect: 10*  
           *Word Size: 3*  
 30       *Filter: on*

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least

150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain  
5 DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

10 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the  
15 stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity.  
20 Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature  
under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic  
25 strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

30 High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance,

sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency  
5 conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C<sub>0</sub>t or R<sub>0</sub>t analysis) or formed between one  
10 nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

15 "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TFZN which is  
20 capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TFZN which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

25 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of TFZN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TFZN.

30 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a

functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

5       “Peptide nucleic acid” (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

10       “Post-translational modification” of an TFZN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TFZN.

      “Probe” refers to nucleic acid sequences encoding TFZN, their complements, or fragments  
15 thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. “Primers” are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target  
20 DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

      Probes and primers as used in the present invention typically comprise at least 15 contiguous  
nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100,  
25 or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

      Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold  
30 Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge

MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

5        “Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

10        An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

15        The term “sample” is used in its broadest sense. A sample suspected of containing TFZN, nucleic acids encoding TFZN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

20        The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

25        The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

30        “Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A “transcript image” refers to the collective pattern of gene expression by a particular cell

type or tissue under given conditions at a given time.

“Transformation” describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term “transformed cells” includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A “transgenic organism,” as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), *supra*.

A “variant” of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an “allelic” (as defined above), “splice,” “species,” or “polymorphic” variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to



another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass “single nucleotide polymorphisms” (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A “variant” of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the “BLAST 2 Sequences” tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## THE INVENTION

The invention is based on the discovery of new human transcription factors and zinc finger proteins (TFZN), the polynucleotides encoding TFZN, and the use of these compositions for the diagnosis, treatment, or prevention of cell proliferative disorders, including cancer, developmental, autoimmune/inflammatory and neurological disorders.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are transcription factors and zinc finger proteins. For example, SEQ ID NO:2 has 36% local identity to *Arabidopsis thaliana* RING-H2 finger protein (GenBank ID g3790573) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.0e-12, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:2 also contains a C3HC4 type (RING finger) domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) The presence of this motif is confirmed by comparison to the DOMO database of protein domains, providing further corroborative evidence that SEQ ID NO:3 is a RING finger protein.

SEQ ID NO:3 is 73% identical to human repressor transcriptional factor, a DNA binding protein related to the ZnF-91 gene family (GenBank ID g1017722) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 1.0e-224, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 also contains Zinc finger, C2H2 type domains and a KRAB box domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:3 is a zinc finger protein.

SEQ ID NO:5 is 68% identical to human nuclear respiratory factor-2 subunit beta 1 (GenBank ID g531895), a GA (i.e. purine rich) binding protein (GABP), as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 5.9e-132, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 is also 64% identical to a murine GABP (GenBank ID g567202), based on BLAST analysis, with a probability score of 1.9e-112. SEQ ID NO:5 also contains ankyrin repeat sequences as determined by searching for statistically significant matches in the hidden Markov model (HMM)-

based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS analysis provide further corroborative evidence for the presence of ankyrin repeat sequences.

SEQ ID NO:6 is 44% identical to a Drosophila Kruppel protein homologue (GenBank ID g5565928), as determined by BLAST analysis. The BLAST probability score is  $2.1e-47$ . Kruppel is a sequence-specific, DNA-binding protein that represses RNA polymerase II transcription in cultured cells of both Drosophila and mammalian origin.

SEQ ID NO:7 is 82% identical to the murine, developmentally-regulated, zinc finger repressor protein, NK10 (GenBank ID g506502), as determined by BLAST analysis, with a probability score of  $5.8e-299$ . SEQ ID NO:7 also contains multiple zinc finger domains and a KRAB (Kruppel-associated box) domain, as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. A KRAB domain is a 75-amino acid transcriptional repressor motif found in eukaryotic zinc finger proteins. Data from BLIMPS and MOTIFS analyses provide further corroborative evidence that SEQ ID NO:7 is a zinc finger protein.

SEQ ID NO:8 is 35% identical to a human homologue of a Drosophila trithorax polypeptide (GenBank ID g2358287). The BLAST probability score is  $6.0e-17$ . Trithorax group polypeptides, in concert with the Polycomb group polypeptides, are both positive and negative transcriptional regulators that determine body structure morphology during Drosophila development. SEQ ID NO:1 and SEQ ID NO:4 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-8 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:9-16 or that distinguish between SEQ ID NO:9-16 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 2397546F6 is the identification number of an Incyte cDNA sequence, and THP1AZT01 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71938382V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g6197637) which contributed to the assembly of the full length polynucleotide sequences. In addition, the identification numbers in column 5 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation “ENST”). Alternatively, the identification numbers in column 5 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation “NM” or “NT”) or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation “NP”). Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon stitching” algorithm. For example, FL\_XXXXXX\_N<sub>1</sub>\_N<sub>2</sub>\_YYYYY\_N<sub>3</sub>\_N<sub>4</sub> represents a “stitched” sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYYY is the number of the prediction generated by the algorithm, and N<sub>1,2,3...</sub>, if present, represent specific exons that may have been manually edited during analysis (See Example V). To illustrate, GNN:g4156137\_004 is the identification number of a Genscan-predicted coding sequence, with g4156137 being the GenBank identification number of the sequence to which Genscan was applied. Alternatively, the identification numbers in column 5 may refer to assemblages of exons brought together by an “exon-stretching” algorithm. For example, FLXXXXXX\_gAAAAA\_gBBBBB\_1\_N is the identification number of a “stretched” sequence, with XXXXXX being the Incyte project identification number, gAAAAA being the GenBank identification number of the human genomic sequence to which the “exon-stretching” algorithm was applied, gBBBBB being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and N referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the “exon-stretching” algorithm, a RefSeq identifier (denoted by “NM,” “NP,” or “NT”) may be used in place of the GenBank identifier (*i.e.*, gBBBBB).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs
GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TFZN variants. A preferred TFZN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TFZN amino acid sequence, and which contains at least one functional or structural characteristic of TFZN.

The invention also encompasses polynucleotides which encode TFZN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-16, which encodes TFZN. The polynucleotide sequences of SEQ ID NO:9-16, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TFZN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TFZN. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:9-

16 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:9-16. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TFZN.

5 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TFZN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in  
10 accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TFZN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode TFZN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TFZN under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TFZN or  
15 its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TFZN and its derivatives without altering the encoded amino acid sequences  
20 include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode TFZN and TFZN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems  
25 using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TFZN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:9-16 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and  
30 S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied

Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV),  
5 PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M.  
10 (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding TFZN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed,  
15 restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids  
20 Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences  
25 are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National  
30 Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include

sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TFZN may be cloned in recombinant DNA molecules that direct expression of TFZN, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TFZN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TFZN-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TFZN, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular



evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TFZN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) *Nucleic Acids Symp. Ser. 7*:215-223; and Horn, T. et al. (1980) *Nucleic Acids Symp. Ser. 7*:225-232.) Alternatively, TFZN itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TFZN, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TFZN, the nucleotide sequences encoding TFZN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TFZN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding TFZN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TFZN and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers

appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding TFZN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TFZN. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; and Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5(6):350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90(13):6340-6344; Buller, R.M. et al. (1985) *Nature* 317(6040):813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31(3):219-226; and Verma, I.M. and N. Somia (1997) *Nature* 389:239-242.) The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding TFZN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TFZN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Life Technologies). Ligation of sequences encoding TFZN into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of

nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TFZN are needed, e.g. for the production of antibodies, vectors which direct high level expression of TFZN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

5 Yeast expression systems may be used for production of TFZN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra;  
10 Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of TFZN. Transcription of sequences encoding TFZN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J.  
15 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill,  
20 New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TFZN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain  
25 infective virus which expresses TFZN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of  
30 DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of

TFZN in cell lines is preferred. For example, sequences encoding TFZN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before  
5 being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These  
10 include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk<sup>r</sup>* and *apr<sup>r</sup>* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat*  
15 confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins  
20 (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest  
25 is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TFZN is inserted within a marker gene sequence, transformed cells containing sequences encoding TFZN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TFZN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates  
30 expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TFZN and that express TFZN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or

chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TFZN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and  
5 fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TFZN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-  
10 Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TFZN include  
15 oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding TFZN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety  
20 of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TFZN may be cultured under  
25 conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TFZN may be designed to contain signal sequences which direct secretion of TFZN through a prokaryotic or eukaryotic cell membrane.

30 In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells

which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

5 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TFZN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric TFZN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TFZN activity. Heterologous protein and  
10 peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins,  
15 respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TFZN encoding sequence and the heterologous protein sequence, so that TFZN may be cleaved away from the heterologous moiety following purification. Methods for fusion protein  
20 expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TFZN may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the  
25 T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

TFZN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TFZN. At least one and up to a plurality of test compounds may be screened for specific binding to TFZN. Examples of test compounds include antibodies, oligonucleotides,  
30 proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TFZN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TFZN

binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TFZN, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or E. coli. Cells expressing TFZN or cell membrane fractions which contain TFZN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TFZN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TFZN, either in solution or affixed to a solid support, and detecting the binding of TFZN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

TFZN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TFZN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TFZN activity, wherein TFZN is combined with at least one test compound, and the activity of TFZN in the presence of a test compound is compared with the activity of TFZN in the absence of the test compound. A change in the activity of TFZN in the presence of the test compound is indicative of a compound that modulates the activity of TFZN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TFZN under conditions suitable for TFZN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TFZN may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding TFZN or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP

system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330).

Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding TFZN may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding TFZN can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TFZN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TFZN, e.g., by secreting TFZN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## **THERAPEUTICS**

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TFZN and transcription factors and zinc finger proteins. In addition, the expression of TFZN is closely associated with adrenal and kidney tissues and a variety of disease states that involve inappropriate cell proliferation, including cancer and Patau's syndrome. Therefore, TFZN appears to play a role in cell proliferative disorders, including cancer, and developmental, autoimmune/inflammatory and neurological disorders. In the treatment of disorders associated with increased TFZN expression or activity, it is desirable to decrease the expression or activity of TFZN. In the treatment of disorders associated with decreased TFZN expression or activity, it is desirable to increase the expression or activity of TFZN.

Therefore, in one embodiment, TFZN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TFZN. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary



thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, a seizure disorder such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a neurological disorder such as Alzheimer's disease, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy,

neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; 5 seasonal affective disorder (SAD); akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia.

In another embodiment, a vector capable of expressing TFZN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased 10 expression or activity of TFZN including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TFZN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TFZN including, but not limited to, those provided above.

15 In still another embodiment, an agonist which modulates the activity of TFZN may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TFZN including, but not limited to, those listed above.

In a further embodiment, an antagonist of TFZN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TFZN. Examples of such 20 disorders include, but are not limited to, those cell proliferative disorders, including cancer, and developmental, autoimmune/inflammatory and neurological disorders described above. In one aspect, an antibody which specifically binds TFZN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express TFZN.

25 In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TFZN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TFZN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate 30 therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TFZN may be produced using methods which are generally known in the art. In particular, purified TFZN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TFZN. Antibodies to TFZN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TFZN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TFZN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TFZN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TFZN may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TFZN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton,

D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TFZN may also be generated. For example, such fragments include, but are not limited to,  $F(ab')_2$  fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the  $F(ab')_2$  fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TFZN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TFZN epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TFZN. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of TFZN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TFZN epitopes, represents the average affinity, or avidity, of the antibodies for TFZN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular TFZN epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the TFZN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TFZN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine

the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TFZN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for  
5 antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TFZN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA,  
10 RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding TFZN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TFZN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense  
15 sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral  
20 vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.*  
25 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TFZN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined  
30 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal,

R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TFZN expression or regulation causes disease, the expression of TFZN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TFZN are treated by constructing mammalian expression vectors encoding TFZN and introducing these vectors by mechanical means into TFZN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TFZN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TFZN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra)), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TFZN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental

parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

5           In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TFZN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TFZN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences  
10 required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al.  
15 (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference.  
20 Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4<sup>+</sup> T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

25           In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TFZN to cells which have one or more genetic abnormalities with respect to the expression of TFZN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas  
30 (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding TFZN to target cells which have one or more genetic abnormalities with respect to the expression of TFZN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TFZN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TFZN to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for TFZN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TFZN-coding RNAs and the synthesis of high levels of TFZN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of TFZN into a variety of cell types. The specific transduction of a subset of cells in a



population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding TFZN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TFZN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends

of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, 5 guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TFZN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, 10 transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TFZN expression or activity, a compound which specifically inhibits expression of the polynucleotide 15 encoding TFZN may be therapeutically useful, and in the treatment of disorders associated with decreased TFZN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TFZN may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method 20 commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a 25 polynucleotide encoding TFZN is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TFZN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the 30 polynucleotide encoding TFZN. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide

can be carried out, for example, using a Schizosaccharomyces pombe gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library  
5 of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells  
10 taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of  
15 such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various  
20 formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of TFZN, antibodies to TFZN, and mimetics, agonists, antagonists, or inhibitors of TFZN.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal,  
25 intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-  
30 acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TFZN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TFZN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TFZN or fragments thereof, antibodies of TFZN, and agonists, antagonists or inhibitors of TFZN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the  $ED_{50}$  (the dose therapeutically effective in 50% of the population) or  $LD_{50}$  (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the  $LD_{50}/ED_{50}$  ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the  $ED_{50}$  with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1  $\mu\text{g}$  to 100,000  $\mu\text{g}$ , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

In another embodiment, antibodies which specifically bind TFZN may be used for the diagnosis of disorders characterized by expression of TFZN, or in assays to monitor patients being treated with TFZN or agonists, antagonists, or inhibitors of TFZN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TFZN include methods which utilize the antibody and a label to detect TFZN in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TFZN, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of TFZN expression. Normal or standard values for TFZN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TFZN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TFZN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TFZN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of TFZN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TFZN, and to monitor regulation of TFZN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TFZN or closely related molecules may be used to identify nucleic acid sequences which encode TFZN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the

probe identifies only naturally occurring sequences encoding TFZN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TFZN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:9-16 or from genomic sequences including promoters, enhancers, and introns of the TFZN gene.

Means for producing specific hybridization probes for DNAs encoding TFZN include the cloning of polynucleotide sequences encoding TFZN or TFZN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as  $^{32}\text{P}$  or  $^{35}\text{S}$ , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding TFZN may be used for the diagnosis of disorders associated with expression of TFZN. Examples of such disorders include, but are not limited to, a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma; and, in particular, a cancer of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; a developmental disorder such as renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, a seizure disorder such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic

lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a neurological disorder such as Alzheimer's disease, epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia. The polynucleotide sequences encoding TFZN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TFZN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TFZN may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TFZN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard

value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TFZN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TFZN, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TFZN, under conditions suitable for hybridization or amplification.

Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding TFZN may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding TFZN, or a fragment of a polynucleotide complementary to the polynucleotide encoding TFZN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TFZN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are



substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TFZN are used to amplify DNA using the  
5 polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as  
10 DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the  
15 alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TFZN include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from  
standard curves. (See, e.g., Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C.  
20 et al. (1993) Anal. Biochem. 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the  
25 polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor  
30 progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic

profile.

In another embodiment, TFZN, fragments of TFZN, or antibodies specific for TFZN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

5 A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No.  
10 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The  
15 resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention  
20 may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000)  
25 Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested  
30 compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for

example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

5           In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an  
10   untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

          Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome  
15   can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by  
20   isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently  
25   positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial  
30   sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

          A proteomic profile may also be generated using antibodies specific for TFZN to quantify the levels of TFZN expression. In one embodiment, the antibodies are used as elements on a microarray,

and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding TFZN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding TFZN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, TFZN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a

solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TFZN and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TFZN, or fragments thereof, and washed. Bound TFZN is then detected by methods well known in the art. Purified TFZN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TFZN specifically compete with a test compound for binding TFZN. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TFZN.

In additional embodiments, the nucleotide sequences which encode TFZN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/234,903, U.S. Ser. No. 60/244,505 and U.S. Ser. No. 60/254,402, are hereby expressly incorporated by reference.

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol

or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERScript plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, *supra*, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in

384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

### III. Sequencing and Analysis

5 Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared  
10 using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI  
15 protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing  
20 vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family  
25 databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or  
30 Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may



begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:9-16. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative transcription factors and zinc finger proteins were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transcription factors and zinc finger proteins, the encoded polypeptides were analyzed by querying against PFAM models for transcription factors and zinc finger proteins. Potential transcription factors and zinc finger proteins were also identified by homology to Incyte cDNA sequences that had been annotated as transcription factors and zinc finger

proteins. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbprl public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any

5 Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length

10 polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

## **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

### **“Stitched” Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene

15 identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a

20 full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals

25 thus identified were then “stitched” together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared

30 by BLAST analysis to the genpept and gbprl public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

### **“Stretched” Sequences**

Partial DNA sequences were extended to full length with an algorithm based on BLAST

analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

5 Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were  
10 therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

#### **VI. Chromosomal Mapping of TFZN Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:9-16 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other  
15 implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:9-16 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences  
20 had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between  
25 chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site  
30 (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:15 was mapped to chromosome 4 within the interval from 169.10 to 178.40 centiMorgans.

#### **VII. Analysis of Polynucleotide Expression**

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

5 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

10 The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by  
20 gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the  
25 other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding TFZN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA  
30 sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or

urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided  
5 by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TFZN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### **VIII. Extension of TFZN Encoding Polynucleotides**

Full length polynucleotide sequences were also produced by extension of an appropriate  
10 fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target  
15 sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR  
20 was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing  $Mg^{2+}$ ,  $(NH_4)_2SO_4$ , and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN  
30 quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by

electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and  
5 sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction  
10 site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following  
15 parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing  
20 primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotide sequences are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

## 25 IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:9-16 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06  
30 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ -<sup>32</sup>P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing 10<sup>7</sup> counts per minute of the labeled probe is used in a typical membrane-based

hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16  
5 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

#### **X. Microarrays**

10 The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra.), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra.). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a  
15 procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998)  
20 Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the  
25 biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on  
30 the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### **Tissue or Cell Sample Preparation**

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is

reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with

5 GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.

10 (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### **Microarray Preparation**

15 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

20 Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C

25 oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

30 Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.



### **Hybridization**

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65° C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60° C. The arrays are washed for 10 min at 45° C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45° C in a second wash buffer (0.1X SSC), and dried.

### **Detection**

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC

computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

#### 10 XI. Complementary Polynucleotides

Sequences complementary to the TFZN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TFZN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TFZN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TFZN-encoding transcript.

#### XII. Expression of TFZN

20 Expression and purification of TFZN is achieved using bacterial or virus-based expression systems. For expression of TFZN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac (tac)* hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TFZN upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TFZN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TFZN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc.

Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

In most expression systems, TFZN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TFZN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TFZN obtained by these methods can be used directly in the assays shown in Examples XVI, and XVI where applicable.

### **XIII. Functional Assays**

TFZN function is assessed by expressing the sequences encoding TFZN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10  $\mu$ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2  $\mu$ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are

discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TFZN on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding TFZN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding TFZN and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### 10 **XIV. Production of TFZN Specific Antibodies**

TFZN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TFZN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TFZN activity by, for example, binding the peptide or TFZN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### **XV. Purification of Naturally Occurring TFZN Using Specific Antibodies**

Naturally occurring or recombinant TFZN is substantially purified by immunoaffinity chromatography using antibodies specific for TFZN. An immunoaffinity column is constructed by covalently coupling anti-TFZN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing TFZN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of TFZN (e.g., high ionic strength

buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TFZN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TFZN is collected.

#### **XVI. Identification of Molecules Which Interact with TFZN**

5 TFZN, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TFZN, washed, and any wells with labeled TFZN complex are assayed. Data obtained using different concentrations of TFZN are used to calculate values for the number, affinity, and association of TFZN with the  
10 candidate molecules.

Alternatively, molecules interacting with TFZN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) *Nature* 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

TFZN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)  
15 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### **XVII. Demonstration of TFZN Activity**

The following assay can be used to demonstrate either transcription factor activity or zinc  
20 finger activity. TFZN is measured by its ability to stimulate transcription of a reporter gene (Liu, H.Y. et al. (1997) *EMBO J.* 16:5289-5298). The assay entails the use of a well characterized reporter gene construct, LexA<sub>op</sub>-LacZ, that consists of LexA DNA transcriptional control elements (LexA<sub>op</sub>) fused to sequences encoding the *E. coli* LacZ enzyme. The methods for constructing and expressing fusion genes, introducing them into cells, and measuring LacZ enzyme activity, are well known to those  
25 skilled in the art. Sequences encoding TFZN are cloned into a plasmid that directs the synthesis of a fusion protein, LexA-TFZN, consisting of TFZN and a DNA binding domain derived from the LexA transcription factor. The resulting plasmid, encoding a LexA-TFZN fusion protein, is introduced into yeast cells along with a plasmid containing the LexA<sub>op</sub>-LacZ reporter gene. The amount of LacZ enzyme activity associated with LexA-TFZN transfected cells, relative to control cells, is proportional  
30 to the amount of transcription stimulated by the TFZN.  
control cells, is proportional to the amount of transcription stimulated by the TFZN.

Alternatively, Zinc finger activity is measured by its ability to bind zinc. A 5-10 micromolar sample solution in 2.5 mM ammonium acetate solution at pH 7.4 is combined with 0.05 M zinc sulfate solution (Aldrich, Milwaukee WI) in the presence of 100 micromolar dithiothreitol with 10% methanol

added. The sample and zinc sulfate solutions are allowed to incubate for 20 minutes. The reaction solution is passed through a Vydac column with approximately 300 Angstrom bore size and 5 micromolar particle size to isolate zinc-sample complex from the solution, and into a mass spectrometer (PE Sciex, Ontario, Canada). Zinc bound to sample is quantified using the functional  
5 atomic mass of 63.5 Da observed by Whittal, R. M. et al. ((2000) Biochemistry 39:8406-8417).

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be  
10 understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID
2729623	1	2729623CD1	9	2729623CB1
7474993	2	7474993CD1	10	7474993CB1
55022684	3	55022684CD1	11	55022684CB1
70287215	4	70287215CD1	12	70287215CB1
5500054	5	5500054CD1	13	5500054CB1
1349648	6	1349648CD1	14	1349648CB1
3888314	7	3888314CD1	15	3888314CB1
3276670	8	3276670CD1	16	3276670CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank NO:	ID	Probability Score	GenBank Homolog
1	2729623CD1	g6682873		1.40E-126	[Homo sapiens] reduced expression in cancer
2	7474993CD1	g3790573		2.00E-12	[Arabidopsis thaliana] RING-H2 finger protein RHA3a Jensen, R.B. et al. (1998) FEBS Lett 436:283-287
3	55022684CD1	g1017722		1.00E-224	[Homo sapiens] repressor transcriptional factor Poncelet D.A., et al. (1998) DNA Cell Biol. 17:931-943
4	70287215CD1	g1613858		8.50E-175	[Homo sapiens] zinc finger protein zfp47 Petroni, D. et al. (1998) DNA Seq. 9:163-169
5	5500054CD1	g531895		6.00E-132	[Homo sapiens] nuclear respiratory factor-2 subunit beta 1 Gugreja, S. et al. (1995) Mol. Cell. Biol. 15:102-111
6	1349648CD1	g5565928		2.20E-47	[Drosophila melanogaster] Kruppel homolog
7	3888314CD1	g506502		5.90E-299	[Mus musculus] NK10 Lange, R. et al. (1995) DNA Cell Biol. 14:971-981
8	3276670CD1	g2358287		6.10E-17	[Homo sapiens] ALR Prasad, R. et al. (1997) Oncogene 15:549-560



Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	2729623CD1	355	S87 S89 S143 S188 T240 S285 T124 S244	N238	ATP/GTP binding site motif signal peptide: M1-A32 transmembrane domain: V13-Y30, Y54-F74, F171-I194, V210-Y230 DHHC zinc finger domain zf-DHHC: Y117-Y181 C. elegans SRA family integral membrane protein signature PR00697F: V53-F74 YOR034C; MEMBRANE DM05142 Q09701 316-569: Y127-V235	MOTIFS HMMER HMMER
2	7474993CD1	432	T285 S56 T356 S8 S14 S84 S365	N414	Zinc finger, C3HC4 type (RING finger) zf-C3HC4: C380-C420 ZINC FINGER, C3HC4 TYPE DM00063 Q06003 119-171: C380-D423	HMMER_PPFAM BLAST_DOMO
3	55022684CD1	577	S149 S151 S28 S372 S45 S512 S80 T200 T278 T306 T334 T36 T418 T474 T86	N299 N327 N383 N411 N415 N495 N499 N523 N551 N560	Zinc finger, C2H2 type zf-C2H2: Y401-H423 Y429-H451 Y457-H479 Y485-H507 Y513-H535 F205-H227 Y233-H255 F261-H283 Y289-H311 Y317-H339 Y345-H367 Y373-H395 KRAB box KRA:L35-K98 KRAB BOX DOMAIN DM00605 Q05481 10-83: L35-P106 KRAB BOX DOMAIN DM00605 Q03923 1-75: L35-P106 KRAB BOX DOMAIN DM00605 P28160 1-69: D39-P106 KRAB BOX DOMAIN DM00605 S22564 1-63: F44-P106 ZINC FINGER METAL BINDING DNA BINDING PROTEIN ZINC FINGER NUCLEAR TRANSCRIPTION REGULATION REPEAT PD008015: R99-G201	HMMER_PPFAM BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3					PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN FINGER PW1 PD017719: G257-H507	BLAST_PRODROM
					ZINC FINGER METAL BINDING DNA BINDING PROTEIN FINGER ZINC NUCLEAR REPEAT TRANSCRIPTION REGULATION PD001562: L35-K98	BLAST_PRODROM
					ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K455-C518	BLAST_PRODROM
					Zinc finger, C2H2 type BL00028: C291-H307	BLIMPS_BLOCKS
					C2H2-type zinc finger signature PR00048: P456-S469	BLIMPS_PRINTS
					PROTEIN ZINC-FINGER METAL BINDING PD000066: H391-C403	BLIMPS_PRODROM
					PROTEIN ZINC FINGER ZINC PD01066: F37-A75	BLIMPS_PRODROM
					Zinc_Finger_C2H2 type, domain: C207-H227 C235-H255 C263-H283 C291-H311 C319-H339 C347-H367 C375-H395 C403-H423 C431-H451 C459-H479 C487-H507 C515-H535	MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	70287215CD1	510	S113 S142 S15 S36 S384 S411 S437 S441 S451 T200 T262 T9	N279	Zinc finger, C2H2 type zf-C2H2: Y480-H502 H286-H308 Y314-H336 Y342-H364 Y370-H392 Y398-H420 Y452-H474 SCAN domain: L40-V135 ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 789-829: Q333-D374, R305-E346, Q361-C400 ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 Q05481 831-885: C291-E346 ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 P52743 31-93: L329-H392 ZINC FINGER, C2H2 TYPE, DOMAIN DM00002 P08042 314-358: C347-H392 ZINC FINGER METAL BINDING ZINC FINGER PROTEIN DNA BINDING NUCLEAR TRANSCRIPTION REGULATION REPEAT PD004640: M1-V135 PROTEIN ZINC FINGER METAL BINDING DNA BINDING ZINC FINGER PATERNALLY EXPRESSED ZN FINGER PW1 PD017719: K278-V503 ZINC FINGER DNA BINDING PROTEIN METAL BINDING NUCLEAR ZINC FINGER TRANSCRIPTION REGULATION REPEAT PD000072: K312-C375 P373C6.1 ZINC FINGER METAL BINDING DNA BINDING PD164437: C251-H286 C2H2-type zinc finger signature PR00048: P341-S354, L357-G366 Zinc finger, C2H2 type, domain proteins: BL00028 Zinc_Finger_C2H2 type, domain C288-H308 C316-H336 C344-H364 C372-H392 C400-H420 C454-H474 C482-H502	HMMER_PFAM  HMMER_PFAM BLAST_DOMO  BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_PRODROM BLAST_PRODROM BLAST_PRODROM BLAST_PRODROM BLIMPS_PRINTS BLIMPS_BLOCKS MOTIFS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	5500054CD1	448	S2, S52, T68, T112, S134, S139, T187, T206, S241, T301, T333, S344, Y385, T403, T417	N219	TRANSCRIPTION; GA; GABP [GA (purine-rich) binding domain] sequence binding protein: DM05409 A53950 126-365: S199-D401, Y126-E328, G94-L123 SUBUNIT NUCLEAR GA (purine-rich) BINDING PROTEIN CHAIN TRANSCRIPTION REGULATION ANKYRIN REPEAT: PD007609: F136-L357; PD008006: M1-L37 REPEAT PROTEIN, ANKYRIN NUCLEOTIDE: PD00078: D35-Q47 Ankyrin repeat: L37-K69, V70-M102, L103-K135	BLAST-DOMO  BLAST-PRODROM  BLIMPS-PRODROM HMMER-PFAM
6	1349648CD1	247	T20, T127, S144, T193, S197, S198		KRAB box KRA: V14-E76 Zinc finger, C2H2 type zf-C2H2: Y211-H233, H254-H276, F282-H304, Y310-H332, Y338-H360, F366-H388, F394-H416, Y450-H472, Y497-H519, Y525-H547, Y553-H575, Y581-H603, Y609-H631 Zinc finger, C2H2 type: C213-H233 C256-H276 C284-H304 C312-H332 C340-H360 C368-H388 C396-H416 C452-H472 C499-H519 C527-H547 C555-H575 C583-H603 C611-H631 Zinc finger, C2H2 type: BL00028: C611-H627 C2H2 type Zinc finger signature: PR00048: P365-C378, L269-G278,	HMMER-PFAM HMMER-PFAM  MOTIFS  BLIMPS-BLOCKS BLIMPS-PRINTS
7	3888314CD1	636	T15, T24, S38, S75, T89, T116, S125, T144, T150, S183, T227, T232, S346, T354, T438, T443, T510, S511, S535, S539, T541, T594, S610	N46, N241, N426, N617	ZINC FINGER PROTEIN 90 ZFP90 NK10 ZINC FINGER METAL-BINDING DNA-BINDING NUCLEAR REPEAT TRANSCRIPTION REGULATOR: PD075871: E78-K209, G390-H631 KRAB domain: DM00605 48689 11-85: Q11-P85	BLAST-PRODROM  BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	3276670CD1	1603	S17, T34, S46, S70, T71, S106, S107, S176, S179, S190, T252, T319, T387, S447, S462, T463, T492, S608, T639, S641, S686, S690, T735, S744, T804, S811, S842, S860, S864, S868, S916, T988, S1039, T1162, S1166, S1190, S1199, S1228, S1262, T1266, S1281, T1353, T1388, S1445, S1460, T1511	N489, N606, N747, N1220	Leucine zipper (leucine_zipper): L124-L145, L131-L153	MOTIFS

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
9	2729623CB1	1629	1518-1629	2729623F6 (OVAR1UT04)	639	1208
				7459387H1 (LIVRTUE01)	1	526
				8112368H1 (OSTEUNC01)	1091	1629
				7111074H1 (SINTFEE01)	500	937
10	7474993CB1	3961	1-408, 1146-1177, 3344-3455, 2471-2605	6755545J1 (SINTFER02)	1821	2434
				7743387H1 (ADRETUE04)	744	1370
				7708089H1 (PANCNOE02)	133	552
				381580R6 (HYPONOB01)	2920	3465
				7619782H1 (KIDNTUE01)	2235	2893
				4292191H1 (BRABDIR01)	3544	3793
				6479189H1 (PROSTMC01)	2710	3329
				2774055H1 (PANCNOT15)	3703	3961
				7621390J1 (HEARFEE03)	1419	2157
				7408851H1 (BRAIFEJ02)	207	783
11	55022684CB1	2509	2291-2509, 1922-2195	2194773F6 (THYRTUT03)	3288	3776
				96197637	1	350
				2008724R6 (TESTNOT03)	1263	1869
				2397546F6 (THPLAZT01)	1627	2154
				7005130H1 (COLNFEC01)	1	568
				5902943F6 (BRAYDIN03)	1790	2509
12	70287215CB1	2119	1-522, 989-1198	FL55022684_g7582558_g1 86774	228	2014
				71938382V1	1041	1604
				55115167H1	1	854
				71204488V1	1587	2113
				GNN:g4156137_004	587	2119
				72009876V1	595	1091
13	5500054CB1	1791	812-922, 158-187, 1156-1317	6559615F6 (BRAFNON02)	1	595
				6559615H1 (BRAFNON02)	1	666
				7639379H1 (SEMVTDE01)	47	695
				040127H1 (TBLYNOT01)	59	375
				8012859H1 (HEARNOC04)	158	473
				4825328F8 (BLADDIT01)	187	646
				g274490	188	468
				GBI.g9053126_regenscan	190	925

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
13				3417548H1 (PTHYNOT04)	247	484
				3417548F7 (PTHYNOT04)	247	813
				8012520H1 (HEARNOC04)	274	846
				7663538J1 (UTRSTME01)	306	908
				5614862H1 (THYMNOR02)	430	694
				5614862F8 (THYMNOR02)	430	886
				5614862F6 (THYMNOR02)	431	969
				5499654H1 (BRABDIR01)	664	888
				5500054H1 (BRABDIR01)	664	904
				5499654F6 (BRABDIR01)	664	904
				6243835H1 (TESTNOT17)	777	1361
				6247981H1 (TESTNOT17)	805	1361
				8181926H1 (EYERNON01)	833	1447
				4072949H1 (KIDNNOT26)	841	955
				4073607H1 (KIDNNOT26)	841	1130
				4073607F6 (KIDNNOT26)	841	1314
				4072949F8 (KIDNNOT26)	841	1354
				4073607T6 (KIDNNOT26)	952	1791
				4069650T6 (KIDNNOT26)	952	1791
				7236743H1 (BRAINOY02)	1075	1670
14	1349648CB1	1343	1-27	7322665H1 (BRAFTUE01)	1084	1548
				4072949T8 (KIDNNOT26)	1123	1744
				2918328H1 (THYMFET03)	1225	1514
				2918328F6 (THYMFET03)	1225	1777
				2918328T6 (THYMFET03)	1225	1791
				7309639U1	1226	1298
				4248439H1 (BRABDIT01)	1234	1496
				6871394H1 (BRAGNON02)	1285	1791
				871245H1 (LUNGAST01)	1	244
				821429H1 (KERANOT02)	1	254
				260408H1 (HNT2RAT01)	1	295
				g1154468	1	528
				821429R1 (KERANOT02)	1	576
				g1147141	1	607
				4545794H1 (COLXTDT01)	11	272
				2617980H1 (GBLADIT03)	14	228

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
14				2614120H1 (GBLADIT03)	14	268
				2617980F6 (GBLADIT03)	14	370
				3108316H1 (BRSTTUT15)	15	92
				3296145H1 (TLYJINT01)	17	264
				3651217H1 (ENDINOT01)	21	330
				396538H1 (PITUNOT02)	26	304
				396538R1 (PITUNOT02)	26	538
				5172882H1 (EPIBEXT01)	27	280
				789202H1 (PROSTUT03)	45	272
				2512489H1 (LIVRTUT04)	45	281
				789202R1 (PROSTUT03)	45	547
				496158H1 (HNT2NOT01)	47	301
				g1813040	48	299
				3421023H1 (UCMCNOT04)	52	295
				5541078H1 (KIDNPEC01)	86	281
				5988040H1 (MCCLDXT02)	188	461
				g876705	237	702
				g900287	237	769
				g880051	237	858
				3692427H1 (HEAANOT01)	310	598
				2987289H1 (CARGDIT01)	312	474
				g2013318	312	492
				4551306H1 (HELAUNT01)	312	541
				1347755H1 (PROSNOT11)	399	620
				g1897620	421	897
				g2105994	480	1019
				340321R6 (NEUTFMT01)	486	580
				340321H1 (NEUTFMT01)	486	722
				4712281H1 (BRAHCT01)	518	765
				2729348H1 (OVARUT05)	548	798
				4829296H1 (BRAVXT03)	568	833
				4637234H1 (MYEPTXT01)	606	865
				4723629H1 (COLCTUT02)	668	902
				2617980T6 (GBLADIT03)	740	1330
				3601005H1 (DRGINOT01)	742	1048
				1647311H1 (PROSTUT09)	758	962



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
14				2227657H1 (SEWVNOT01)	765	1002
				5529948H1 (KIDNNOT34)	771	1054
				2478073H1 (SMCANOT01)	786	1018
				4186623H1 (BRSTNOT31)	805	1153
				340321R6 (NEUTFMT01)	827	968
				824481H1 (PROSNOT06)	854	1049
				3148271T7 (ADRENON04)	856	1319
				153207H1 (THP1PLB02)	906	1130
				g1156757	907	1343
				1349648H1 (LATRTUT02)	920	1172
				1349648F1 (LATRTUT02)	920	1270
				1343157T6 (COLNTT03)	920	1305
				1349648F6 (LATRTUT02)	920	1343
				2239760H1 (PANCCT02)	957	1212
15	3888314CB1	2489	488-1065, 1-32, 2182-2489	3992986H1 (LUNGNON03)	975	1264
				3839589H1 (DENDTNT01)	1116	1343
				5281793H1 (TESTNON04)	1125	1336
				3840238H1 (DENDTNT01)	1165	1343
				3148271H1 (ADRENON04)	1183	1343
				4698811H1 (BRALNOT01)	1188	1336
				4901405H1 (OVARDT01)	1209	1336
				6549845H1 (BRAFNON02)	1	407
				6549845F8 (BRAFNON02)	1	692
				5532445F8 (HEARFET05)	178	487
				1528228H1 (UCMCL5T01)	197	367
				7273596H1 (PROSUNJ01)	269	899
				7273596F8 (PROSUNJ01)	269	931
				70488217V1	297	487
				g2046867	364	895
				8214012H1 (FIBRTXC01)	448	487
				4140846H1 (BRSTTMT01)	453	735
				4140846F9 (BRSTTMT01)	453	1036
				8214012H1 (FIBRTXC01)	485	1076
				4030602F8 (BRAINOT23)	551	1170
				7273596R8 (PROSUNJ01)	564	1219
				3929925H1 (KIDNNOT19)	630	907

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
15				7073510R6 (BRAUTDR04)	686	1308
				7073510R8 (BRAUTDR04)	686	1336
				111294R1 (PTUNOT01)	899	1349
				111294R6 (PTUNOT01)	899	1424
				3766051F6 (BRSTNOT24)	906	945
				3766451H1 (BRSTNOT24)	906	946
				3765992F6 (BRSTNOT24)	906	946
				3752010F6 (UTRSNOT18)	906	946
				2495993H1 (ADRETUT05)	952	1192
				3692321F6 (HEAANOT01)	953	1029
				3766451H1 (BRSTNOT24)	953	1181
				3766051F6 (BRSTNOT24)	953	1354
				3752010F6 (UTRSNOT18)	953	1368
				3765992F6 (BRSTNOT24)	953	1371
				7081835R8 (STOMTMR02)	1009	1808
				7073510F6 (BRAUTDR04)	1224	2039
				7081835F8 (STOMTMR02)	1250	2027
				4140846T9 (BRSTTMT01)	1330	1930
				4326477F6 (TLYMUNT01)	1351	1842
				4326477H1 (TLYMUNT01)	1353	1528
				3888314H1 (UTRSNOT05)	1402	1679
				7073510H1 (BRAUTDR04)	1434	2039
				7081835H1 (STOMTMR02)	1462	2027
				6594866H2 (BONFTXT01)	1554	1793
				2079383H1 (UTRSNOT08)	1651	1913
				6405368H1 (UTREDIT10)	1665	1952
				7254119H1 (BRAMNOA01)	1835	2460
				1353732H1 (LATRTUT02)	1885	2125
				6759433J1 (HEAONOR01)	1912	2468
				1309523H1 (COLNFET02)	2264	2489

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
16	3276670CB1	6236	5846-6236, 5743-5802, 1-2866, 4069-4207	7381492H1 (ENDMUNE01) 8017763J1 (BMARTXE01) 7358722H1 (BRAIFEE05) 7199295R8 (LUNGFER04) 6777974H1 (OVARDIR01) g1157098 7102781H1 (BRAWTDR02) 7604094H1 (ESOGTME01) 3760513H1 (BRAHDTI03) 6773271J1 (OVARDIR01) 6274096F8 (BRAIFEN03) 7287313H1 (BRAIFER06) 7740943H1 (THYMN0E01) 7434811H1 (PANC DIR02) 7386169H1 (PROSUNE04) 8026092J2 (ENDMUNE01) 6770258J1 (BRAUNOR01) 7384438H1 (FTUBTUE01) 7604094J1 (ESOGTME01) 8221257H1 (COLHTUS02) 1233346H1 (LUNGFET03) 7722925H2 (THYRDIE01) 8053538J1 (FTUBTUE01) g907924 6253351H1 (LUNPTUT02) g6397883 g6505307 7740943J1 (THYMN0E01) g946509 g4436127 6773271H1 (OVARDIR01) 6570153H1 (MCLDTXN05) 8219706J2 (SINTFER02) 8018336J1 (BMARTXE01) g4243863 6868539H1 (BRAGNON02)	1 291 530 623 761 950 977 1190 1266 1273 1341 1536 1640 1659 1745 1806 1807 1818 1828 1916 1950 2014 2025 2073 2119 2120 2126 2126 2128 2216 2299 2455 2455 2455 2455 2454 2462 2462 2461 2751 2418 2386 2768 2773 2813 2827 2831 2867	

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
16				g1493380	2455	2926
				g4282357	2455	2931
				g5768492	2455	2944
				g7375590	2455	2951
				6765095J1 (BRAUNOR01)	2455	3041
				g2064243	2456	2781
				g2069500	2458	2783
				g943897	2459	2942
				g945611	2459	2962
				8217829J1 (SINTER02)	2460	3108
				g5672547	2463	2690
				g5810543	2463	2766
				6770607J1 (BRAUNOR01)	2463	2899
				g6398743	2463	2910
				6768686J1 (BRAUNOR01)	2463	3037
				g898468	2465	2961
				8096381H1 (EYERNOA01)	2484	3210
				g3034177	2508	2637
				g2262018	2520	2917
				7386513H1 (PROSUNE04)	2627	2840
				7386512H1 (PROSUNE04)	2627	2848
				8188924H1 (BMARTXN03)	2757	3355
				55034309H1 (PHOSDNV10)	2819	3449
				6450951H1 (BRAINOC01)	2884	3402
				6450951F8 (BRAINOC01)	2897	3533
				g946508	2958	3291
				g946173	2958	3319
				4252279H1 (BRADDIR01)	2975	3230
				4252279F6 (BRADDIR01)	2975	3596
				6894980J1 (BRAITDR03)	3001	3590
				7744383J1 (ADRETFUE04)	3006	3639
				3366744H1 (CONNTUT04)	3037	3296
				6894980H1 (BRAITDR03)	3044	3617
				8219706H1 (SINTER02)	3047	3685
				6980501H1 (BRAHTDR04)	3135	3581
				3276670F6 (PROSBPT06)	3151	3702

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
16				g943006	3179	3625
				6205278H1 (PITUNON01)	3190	3688
				7733380J2 (COLDDIE01)	3203	3831
				g945610	3214	3625
				8045548H1 (OVRTUE01)	3289	3637
				8045548J1 (OVRTUE01)	3294	3637
				7280536H1 (BMARTXE01)	3332	3880
				3747210F6 (THYMNOT08)	3405	3880
				6768686H1 (BRAUNOR01)	3427	3772
				3276670H1 (PROSBPT06)	3453	3701
				7357424H1 (HEARNON03)	3534	3969
				3747210H1 (THYMNOT08)	3586	3880
				7974017H2 (CONRTUC01)	3592	3950
				55034309J1 (PHOSDNV10)	3632	4267
				6765095H1 (BRAUNOR01)	3642	4281
				6337945H1 (BRANDIN01)	3688	4073
				3329367H1 (HEAONOT04)	3700	3962
				3746723H1 (THYMNOT08)	3700	3991
				8223335H1 (COLHTUS02)	3721	4343
				7717329H1 (SINTFEE02)	3729	4371
				6813256J1 (ADRETUR01)	3732	4398
				6340838H1 (BRANDIN01)	3771	4073
				7732670H2 (COLDDIE01)	3802	4462
				1311978H1 (COLNFET02)	3815	4064
				6862135H1 (BRAGNON02)	3957	4076
				7755566H1 (SPLINTUE01)	4156	4551
				7755566J1 (SPLINTUE01)	4156	4551
				4922525H1 (TESTNOT11)	4204	4475
				7716529J1 (SINTFEE02)	4209	4894
				3765696H1 (BRSTNOT24)	4214	4501
				g694878	4223	4712
				2851486H1 (BRSTTUT13)	4242	4554
				55096551J1 (ADMEDRV02)	4266	4415
				55096551H1 (ADMEDRV02)	4266	4415
				7717329J1 (SINTFEE02)	4303	4894
				7634852J1 (SINTDIE01)	4318	4851

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
16				72044739V1	4350	4439
				72044428V1	4350	4439
				71486736V1	4350	4538
				2020657H1 (CONNNOT01)	4350	4621
				72043675V1	4350	4788
				2020657F6 (CONNNOT01)	4350	4846
				72044425V1	4350	4849
				72043763V1	4350	4947
				72044162V1	4350	4956
				71505060V1	4350	4956
				71502929V1	4353	4664
				71505986V1	4373	4942
				8227483H1 (BRAUTDR02)	4377	4992
				72043066V1	4428	4845
				1422414H1 (KIDNNOT09)	4445	4555
				g1757671	4468	4914
				7608776H1 (COLRTUE01)	4554	4853
				7965131H1 (SPLNFEA02)	4556	4811
				g4450721	4577	4773
				72043669V1	4585	4650
				3639047H1 (LUNGNOT30)	4588	4885
				2859739H1 (SININOT03)	4606	4869
				6813256H1 (ADRETUR01)	4609	5181
				7634852H1 (SINTDIE01)	4634	5160
				72044463V1	4644	5153
				72044283V1	4645	5304
				71503777V1	4647	4916
				7428534H1 (UTRMTMR02)	4647	5161
				72043231V1	4649	5085
				7649054H2 (STOMTDE01)	4667	5167
				7649054J2 (STOMTDE01)	4667	5167
				1870954H1 (SKINBIT01)	4677	4930
				1868933H1 (SKINBIT01)	4677	4956
				1870954F6 (SKINBIT01)	4677	5223
				71502735V1	4679	5098
				5678942H1 (BRAENOT02)	4706	4907

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
16				4618136H1 (BRAYDIT01)	4715	4984
				6344420H1 (LUNGDIS03)	4726	4967
				72043551V1	4733	5048
				72043065V1	4733	5209
				6078686H1 (UTREDIT09)	4736	4907
				7732670J2 (COLDDIE01)	4742	5436
				7642921J1 (SEMTDE01)	4749	5388
				8227483J1 (BRAUTDR02)	4761	5470
				3337240H1 (SPLNNOT10)	4775	5027
				5976047H1 (BRAZNOT01)	4792	5289
				72043714V1	4809	5232
				72044309V1	4810	5147
				7011789H1 (KIDNNOC01)	4816	5368
				72044686V1	4817	5364
				72043835V1	4822	4975
				8044628J1 (OVARUE01)	4825	4966
				72043595V1	4827	5356
				72043194V1	4844	5217
				72044434V1	4851	5293
				1929059R6 (BRSTNOT02)	4867	5357
				7733380H2 (COLDDIE01)	4868	5437
				1929059H1 (BRSTNOT02)	4873	5148
				5903151H1 (BRAYDIN03)	4888	5183
				1570035H1 (UTRSNOT05)	4945	5171
				8087989H1 (BLADTUN02)	4979	5651
				7642921H1 (SEMTDE01)	4984	5634
				72044001V1	5017	5643
				72043383V1	5026	5266
				6911615J1 (PTUDIR01)	5031	5682
				7608776J1 (COLRTUE01)	5039	5600
				5548161H1 (LJNGNOT39)	5042	5310
				3704368H1 (PENCNOT07)	5046	5332
				72044227V1	5052	5553
				72043956V1	5102	5555
				8256350H1 (BRAHDIT10)	5103	5731
				7716004H1 (SINTFEE02)	5123	5598

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment (s)	Sequence Fragments	5' Position	3' Position
16				7958763J1 (TONSDIE01)	5133	5301
				7958763H1 (TONSDIE01)	5135	5301
				6911615H1 (PTUDIR01)	5150	5746
				72044451V1	5159	5658
				72044756V1	5172	5648
				8044628H1 (OVARTUE01)	5183	5759
				72043177V1	5186	5476
				1239423H1 (LUNGUT02)	5190	5418
				8287343T1 (OVARIN02)	5197	5846
				72044652V1	5228	5563
				9564114	5253	5458
				4031763H1 (BRAINOT23)	5260	5530
				2939052H1 (THYMFET02)	5276	5506
				6197370H1 (PTUNON01)	5307	5854
				8076124J1 (ADRETUE02)	5352	5757
				4782571H1 (COLNNOT39)	5360	5610
				72043093V1	5366	5677
				71507250V1	5393	5765
				3998918H1 (HNT2AZS07)	5407	5707
				72042830V1	5413	5566
				71505141V1	5422	5517
				71506964V1	5429	5763
				6144386H1 (BRANDIT03)	5442	5749
				7259095H1 (BRANWOC01)	5453	5914
				1700430H1 (BLADTUT05)	5459	5648
				1699976H1 (BLADTUT05)	5459	5683
				1700430F6 (BLADTUT05)	5461	5904
				7439024H1 (ADRETUE02)	5462	5889
				71502525V1	5488	5747
				3856525H1 (BRAITUT12)	5492	5786
				3856525F6 (BRAITUT12)	5492	5979
				7235443H1 (BRAINOY02)	5503	5916
				708661H1 (SYNORAT04)	5509	5742
				5598063H1 (UTRENON03)	5509	5777
				2414491H1 (HNT3AZT01)	5526	5742
				5772223H1 (BRAINOT20)	5582	5772



Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
16				6852390H1 (BRAIFEN08)	5590	5774
				6852241H1 (BRAIFEN08)	5595	5754
				g1392269	5648	6236
				921931H1	5719	6051
				g6035371	5785	5919
				g5235264	5786	6104
				g3231991	5804	5911

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
9	2729623CB1	BRAUTDR02
10	7474993CB1	CARDNOT01
11	55022684CB1	ADRENOT07
12	70287215CB1	KIDNNOC01
13	5500054CB1	KIDNNOT26
14	1349648CB1	GELANOT01
15	3888314CB1	BRSTNOT24
16	3276670CB1	BRAUNOR01

Table 6

Library	Vector	Library Description
ADRENOT07	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
BRAUNOR01	pINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased satellitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive cells.
BRAUTDR02	PCDNA2.1	This random primed library was constructed using RNA isolated from pooled amygdala and entorhinal cortex tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningial fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver.
BRSTNOT24	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 46-year-old Caucasian female during bilateral subcutaneous mastectomy. Pathology indicated nonproliferative fibrocystic disease. Family history included breast cancer and cardiovascular disease.
CARDNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the cardiac muscle of a 65-year-old Caucasian male, who died from a gunshot wound.
GBLANOT01	pINCY	Library was constructed using RNA isolated from diseased gallbladder tissue removed from a 53-year-old Caucasian female during a cholecystectomy. Pathology indicated mild chronic cholecystitis and cholelithiasis with approximately 150 mixed gallstones. Family history included benign hypertension.
KIDNNOC01	pINCY	This large size-fractionated library was constructed using RNA isolated from pooled left and right kidney tissue removed from a Caucasian male fetus, who died from Patau's syndrome (trisomy 13) at 20-weeks' gestation.

Table 6

Library	Vector	Library Description
KIDNNOT26	pINCY	Library was constructed using RNA isolated from left kidney medulla and cortex tissue removed from a 53-year-old Caucasian female during a nephroureterectomy. Pathology for the associated tumor tissue indicated grade 2 renal cell carcinoma involving the lower pole of the kidney. Patient history included hyperlipidemia, cardiac dysrhythmia, metrorrhagia, normal delivery, cerebrovascular disease, atherosclerotic coronary artery disease, and tobacco abuse. Family history included cerebrovascular disease and atherosclerotic coronary artery disease.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL HDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score $\geq$ GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
  - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-8,
  - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-8,
  - c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, and
  - d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-8.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:9-16.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method of producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

- b) recovering the polypeptide so expressed.

10. A method of claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

5

11. An isolated antibody which specifically binds to a polypeptide of claim 1.

12. An isolated polynucleotide selected from the group consisting of:

- 10 a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16,
- b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:9-16,
- c) a polynucleotide complementary to a polynucleotide of a),
- 15 d) a polynucleotide complementary to a polynucleotide of b), and
- e) an RNA equivalent of a)-d).

13. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 12.

20

14. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:

- 25 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
- b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

30

15. A method of claim 14, wherein the probe comprises at least 60 contiguous nucleotides.

16. A method of detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 12, the method comprising:



- a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and
- b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

5

17. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

18. A composition of claim 17, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

10

19. A method for treating a disease or condition associated with decreased expression of functional TFZN, comprising administering to a patient in need of such treatment the composition of claim 17.

15

20. A method of screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting agonist activity in the sample.

20

21. A composition comprising an agonist compound identified by a method of claim 20 and a pharmaceutically acceptable excipient.

22. A method for treating a disease or condition associated with decreased expression of functional TFZN, comprising administering to a patient in need of such treatment a composition of claim 21.

25

23. A method of screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

30

- a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
- b) detecting antagonist activity in the sample.

24. A composition comprising an antagonist compound identified by a method of claim 23 and a pharmaceutically acceptable excipient.

25. A method for treating a disease or condition associated with overexpression of functional TFZN, comprising administering to a patient in need of such treatment a composition of claim 24.

26. A method of screening for a compound that specifically binds to the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

27. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, the method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

28. A method of screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

29. A method of assessing toxicity of a test compound, the method comprising:

- a) treating a biological sample containing nucleic acids with the test compound,
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising

at least 20 contiguous nucleotides of a polynucleotide of claim 12 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 12 or fragment thereof,

- 5       c)     quantifying the amount of hybridization complex, and  
       d)     comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

10

30. A diagnostic test for a condition or disease associated with the expression of TFZN in a biological sample, the method comprising:

- 15       a)     combining the biological sample with an antibody of claim 11, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex, and  
       b)     detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

31. The antibody of claim 11, wherein the antibody is:

- 20       a)     a chimeric antibody,  
       b)     a single chain antibody,  
       c)     a Fab fragment,  
       d)     a F(ab')<sub>2</sub> fragment, or  
       e)     a humanized antibody.

25

32. A composition comprising an antibody of claim 11 and an acceptable excipient.

33. A method of diagnosing a condition or disease associated with the expression of TFZN in a subject, comprising administering to said subject an effective amount of the composition of claim 32.

30

34. A composition of claim 32, wherein the antibody is labeled.

35. A method of diagnosing a condition or disease associated with the expression of TFZN in a subject, comprising administering to said subject an effective amount of the composition of claim 34.

36. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibodies from said animal, and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

37. A polyclonal antibody produced by a method of claim 36.

38. A composition comprising the polyclonal antibody of claim 37 and a suitable carrier.

39. A method of making a monoclonal antibody with the specificity of the antibody of claim 11, the method comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8, or an immunogenic fragment thereof, under conditions to elicit an antibody response,
- b) isolating antibody producing cells from the animal,
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells,
- d) culturing the hybridoma cells, and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

40. A monoclonal antibody produced by a method of claim 39.

41. A composition comprising the monoclonal antibody of claim 40 and a suitable carrier.

42. The antibody of claim 11, wherein the antibody is produced by screening a Fab expression library.

43. The antibody of claim 11, wherein the antibody is produced by screening a recombinant immunoglobulin library.

44. A method of detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 in the sample.

45. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8 from a sample, the method comprising:

- a) incubating the antibody of claim 11 with a sample under conditions to allow specific binding of the antibody and the polypeptide, and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-8.

46. A microarray wherein at least one element of the microarray is a polynucleotide of claim 13.

47. A method of generating a transcript image of a sample which contains polynucleotides, the method comprising:

- a) labeling the polynucleotides of the sample,
- b) contacting the elements of the microarray of claim 46 with the labeled polynucleotides of the sample under conditions suitable for the formation of a hybridization complex, and
- c) quantifying the expression of the polynucleotides in the sample.

48. An array comprising different nucleotide molecules affixed in distinct physical locations on a solid substrate, wherein at least one of said nucleotide molecules comprises a first oligonucleotide or polynucleotide sequence specifically hybridizable with at least 30 contiguous nucleotides of a target polynucleotide, and wherein said target polynucleotide is a polynucleotide of claim 12.

49. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 30 contiguous nucleotides of said target polynucleotide.

50. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to at least 60 contiguous nucleotides of said target polynucleotide.

51. An array of claim 48, wherein said first oligonucleotide or polynucleotide sequence is completely complementary to said target polynucleotide.

52. An array of claim 48, which is a microarray.

53. An array of claim 48, further comprising said target polynucleotide hybridized to a nucleotide molecule comprising said first oligonucleotide or polynucleotide sequence.

54. An array of claim 48, wherein a linker joins at least one of said nucleotide molecules to said solid substrate.

55. An array of claim 48, wherein each distinct physical location on the substrate contains multiple nucleotide molecules, and the multiple nucleotide molecules at any single distinct physical location have the same sequence, and each distinct physical location on the substrate contains nucleotide molecules having a sequence which differs from the sequence of nucleotide molecules at another distinct physical location on the substrate.

56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

60. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

61. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

62. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

63. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.

5 64. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:9.

65. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:10.

66. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:11.

10

67. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:12.

68. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:13.

15

69. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:14.

70. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:15.

20

71. A polynucleotide of claim 12, comprising the polynucleotide sequence of SEQ ID NO:16.

<110> INCYTE GENOMICS, INC.  
 NGUYEN, Danniell B.  
 YUE, Henry  
 GANDHI, Ameena R.  
 HAFALIA, April J.A.  
 WALIA, Narinder K.  
 YAO, Monique G.  
 THORNTON, Michael  
 RAMKUMAR, Jayalaxmi  
 THANGAVELU, Kavitha  
 LU, Yan  
 LEE, Sally  
 BAUGHN, Mariah R.  
 TANG, Y. Tom  
 AZIMZAI, Yalda  
 KALAFUS, Daniel P.  
 LU, Dyung Aina M.

<120> TRANSCRIPTION FACTORS AND ZINC FINGER PROTEINS

<130> PF-0813 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/234,903; 60/244,505; 60/254,402

<151> 2000-09-22; 2000-10-30; 2000-12-08

<160> 16

<170> PERL Program

<210> 1

<211> 355

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 2729623CD1

<400> 1

Met	Ala	Pro	Trp	Thr	Leu	Trp	Arg	Cys	Cys	Gln	Arg	Val	Val	Gly
1				5					10					15
Trp	Val	Pro	Val	Leu	Phe	Ile	Thr	Phe	Val	Val	Val	Trp	Ser	Tyr
				20					25					30
Tyr	Ala	Tyr	Val	Val	Glu	Leu	Cys	Val	Phe	Thr	Ile	Phe	Gly	Asn
				35					40					45
Glu	Glu	Asn	Gly	Lys	Thr	Val	Val	Tyr	Leu	Val	Ala	Phe	His	Leu
				50					55					60
Phe	Phe	Val	Met	Phe	Val	Trp	Ser	Tyr	Trp	Met	Thr	Ile	Phe	Thr
				65					70					75
Ser	Pro	Ala	Ser	Pro	Ser	Lys	Glu	Phe	Tyr	Leu	Ser	Asn	Ser	Glu
				80					85					90
Lys	Glu	Arg	Tyr	Glu	Lys	Glu	Phe	Ser	Gln	Glu	Arg	Gln	Gln	Glu
				95					100					105
Ile	Leu	Arg	Arg	Ala	Ala	Arg	Ala	Leu	Pro	Ile	Tyr	Thr	Thr	Ser
				110					115					120
Ala	Ser	Lys	Thr	Ile	Arg	Tyr	Cys	Glu	Lys	Cys	Gln	Leu	Ile	Lys



				125					130				135
Pro	Asp	Arg	Ala	His	His	Cys	Ser	Ala	Cys	Asp	Ser	Cys	Ile
				140					145				150
Lys	Met	Asp	His	His	Cys	Pro	Trp	Val	Asn	Asn	Cys	Val	Gly
				155					160				165
Ser	Asn	Tyr	Lys	Phe	Phe	Leu	Leu	Phe	Leu	Leu	Tyr	Ser	Leu
				170					175				180
Tyr	Cys	Leu	Phe	Val	Ala	Ala	Thr	Val	Leu	Glu	Tyr	Phe	Ile
				185					190				195
Phe	Trp	Thr	Asn	Glu	Leu	Thr	Asp	Thr	Arg	Ala	Lys	Phe	His
				200					205				210
Leu	Phe	Leu	Phe	Phe	Val	Ser	Ala	Met	Phe	Phe	Ile	Ser	Val
				215					220				225
Ser	Leu	Phe	Ser	Tyr	His	Cys	Trp	Leu	Val	Gly	Lys	Asn	Arg
				230					235				240
Thr	Ile	Glu	Ser	Phe	Arg	Ala	Pro	Thr	Phe	Ser	Tyr	Gly	Pro
				245					250				255
Gly	Asn	Gly	Phe	Ser	Leu	Gly	Cys	Ser	Lys	Asn	Trp	Arg	Gln
				260					265				270
Phe	Gly	Asp	Glu	Lys	Lys	Tyr	Trp	Leu	Leu	Pro	Ile	Phe	Ser
				275					280				285
Leu	Gly	Asp	Gly	Cys	Ser	Phe	Pro	Thr	Arg	Leu	Val	Gly	Met
				290					295				300
Pro	Glu	Gln	Ala	Ser	Val	Thr	Asn	Gln	Asn	Glu	Tyr	Ala	Arg
				305					310				315
Ser	Gly	Ser	Asn	Gln	Pro	Phe	Pro	Ile	Lys	Pro	Leu	Ser	Glu
				320					325				330
Lys	Asn	Arg	Leu	Leu	Asp	Ser	Glu	Ser	Gln	Trp	Leu	Glu	Asn
				335					340				345
Ala	Glu	Glu	Gly	Ile	Val	Lys	Ser	Gly	Val				
				350					355				

&lt;210&gt; 2

&lt;211&gt; 432

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7474993CD1

&lt;400&gt; 2

Met	Arg	Pro	Trp	Glu	Met	Thr	Ser	Asn	Arg	Gln	Pro	Pro	Ser	Val
1				5					10					15
Arg	Pro	Ser	Gln	His	His	Phe	Ser	Gly	Glu	Arg	Cys	Asn	Thr	Pro
				20					25					30
Ala	Arg	Asn	Arg	Arg	Ser	Pro	Pro	Val	Arg	Arg	Gln	Arg	Gly	Arg
				35					40					45
Arg	Asp	Arg	Leu	Ser	Arg	His	Asn	Ser	Ile	Ser	Gln	Asp	Glu	Asn
				50					55					60
Tyr	His	His	Leu	Pro	Tyr	Ala	Gln	Gln	Gln	Ala	Ile	Glu	Glu	Pro
				65					70					75
Arg	Ala	Phe	His	Pro	Pro	Asn	Val	Ser	Pro	Arg	Leu	Leu	His	Pro
				80					85					90
Ala	Ala	His	Pro	Pro	Gln	Gln	Asn	Ala	Val	Met	Val	Asp	Ile	His
				95					100					105
Asp	Gln	Leu	His	Gln	Gly	Thr	Val	Pro	Val	Ser	Tyr	Thr	Val	Thr
				110					115					120

Thr	Val	Ala	Pro	His	Gly	Ile	Pro	Leu	Cys	Thr	Gly	Gln	His	Ile
				125					130					135
Pro	Ala	Cys	Ser	Thr	Gln	Gln	Val	Pro	Gly	Cys	Ser	Val	Val	Phe
				140					145					150
Ser	Gly	Gln	His	Leu	Pro	Val	Cys	Ser	Val	Pro	Pro	Pro	Met	Leu
				155					160					165
Gln	Ala	Cys	Ser	Val	Gln	His	Leu	Pro	Val	Pro	Tyr	Ala	Ala	Phe
				170					175					180
Pro	Pro	Leu	Ile	Ser	Ser	Asp	Pro	Phe	Leu	Ile	His	Pro	Pro	His
				185					190					195
Leu	Ser	Pro	His	His	Pro	Pro	His	Leu	Pro	Pro	Pro	Gly	Gln	Phe
				200					205					210
Val	Pro	Phe	Gln	Thr	Gln	Gln	Ser	Arg	Ser	Pro	Leu	Gln	Arg	Ile
				215					220					225
Glu	Asn	Glu	Val	Glu	Leu	Leu	Gly	Glu	His	Leu	Pro	Val	Gly	Gly
				230					235					240
Phe	Thr	Tyr	Pro	Pro	Ser	Ala	His	Pro	Pro	Thr	Leu	Pro	Pro	Ser
				245					250					255
Ala	Pro	Leu	Gln	Phe	Leu	Thr	His	Asp	Pro	Leu	His	Gln	Glu	Val
				260					265					270
Ser	Phe	Gly	Val	Pro	Tyr	Pro	Pro	Phe	Met	Pro	Arg	Arg	Leu	Thr
				275					280					285
Gly	Arg	Ser	Arg	Tyr	Arg	Ser	Gln	Gln	Pro	Ile	Pro	Pro	Pro	Pro
				290					295					300
Tyr	His	Pro	Ser	Leu	Leu	Pro	Tyr	Val	Leu	Ser	Met	Leu	Pro	Val
				305					310					315
Pro	Pro	Ala	Val	Gly	Pro	Thr	Phe	Ser	Phe	Glu	Leu	Asp	Val	Glu
				320					325					330
Asp	Gly	Glu	Val	Glu	Asn	Tyr	Glu	Ala	Leu	Leu	Asn	Leu	Ala	Glu
				335					340					345
Arg	Leu	Gly	Glu	Ala	Lys	Pro	Arg	Gly	Leu	Thr	Lys	Ala	Asp	Ile
				350					355					360
Glu	Gln	Leu	Pro	Ser	Tyr	Arg	Phe	Asn	Pro	Asn	Asn	His	Gln	Ser
				365					370					375
Glu	Gln	Thr	Leu	Cys	Val	Val	Cys	Met	Cys	Asp	Phe	Glu	Ser	Arg
				380					385					390
Gln	Leu	Leu	Arg	Val	Leu	Pro	Cys	Asn	His	Glu	Phe	His	Ala	Lys
				395					400					405
Cys	Val	Asp	Lys	Trp	Leu	Lys	Ala	Asn	Arg	Thr	Cys	Pro	Ile	Cys
				410					415					420
Arg	Ala	Asp	Ala	Ser	Glu	Val	His	Arg	Asp	Ser	Glu			
				425					430					

&lt;210&gt; 3

&lt;211&gt; 577

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 55022684CD1

&lt;400&gt; 3

Met	Asp	Asp	Leu	Lys	Tyr	Gly	Val	Tyr	Pro	Leu	Lys	Glu	Ala	Ser
1				5					10					15
Gly	Cys	Pro	Gly	Ala	Glu	Arg	Asn	Leu	Leu	Val	Tyr	Ser	Tyr	Phe
				20					25					30
Glu	Lys	Glu	Thr	Leu	Thr	Phe	Arg	Asp	Val	Ala	Ile	Glu	Phe	Ser

				35					40				45	
Leu	Glu	Glu	Trp	Glu	Cys	Leu	Asn	Pro	Ala	Gln	Gln	Asn	Leu	Tyr
				50					55					60
Met	Asn	Val	Met	Leu	Glu	Asn	Tyr	Lys	Asn	Leu	Val	Phe	Leu	Ala
				65					70					75
Gly	Val	Ala	Val	Ser	Lys	Gln	Asp	Pro	Val	Thr	Cys	Leu	Glu	Gln
				80					85					90
Glu	Lys	Glu	Pro	Trp	Asn	Met	Lys	Arg	His	Glu	Met	Val	Asp	Glu
				95					100					105
Pro	Pro	Ala	Met	Cys	Ser	Tyr	Phe	Thr	Lys	Asp	Leu	Trp	Pro	Glu
				110					115					120
Gln	Asp	Ile	Lys	Asp	Ser	Phe	Gln	Gln	Val	Ile	Leu	Arg	Arg	Tyr
				125					130					135
Gly	Lys	Cys	Glu	His	Glu	Asn	Leu	Gln	Leu	Arg	Lys	Gly	Ser	Ala
				140					145					150
Ser	Val	Asp	Glu	Tyr	Lys	Val	His	Lys	Glu	Gly	Tyr	Asn	Glu	Leu
				155					160					165
Asn	Gln	Cys	Leu	Thr	Thr	Thr	Gln	Ser	Lys	Ile	Phe	Pro	Cys	Asp
				170					175					180
Lys	Tyr	Val	Lys	Val	Phe	His	Lys	Phe	Leu	Asn	Ala	Asn	Arg	His
				185					190					195
Lys	Thr	Arg	His	Thr	Gly	Lys	Lys	Pro	Phe	Lys	Cys	Lys	Lys	Cys
				200					205					210
Gly	Lys	Ser	Phe	Cys	Met	Leu	Leu	His	Leu	Ser	Gln	His	Lys	Arg
				215					220					225
Ile	His	Ile	Arg	Glu	Asn	Ser	Tyr	Gln	Cys	Glu	Glu	Cys	Gly	Lys
				230					235					240
Ala	Phe	Lys	Trp	Phe	Ser	Thr	Leu	Thr	Arg	His	Lys	Arg	Ile	His
				245					250					255
Thr	Gly	Glu	Lys	Pro	Phe	Lys	Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe
				260					265					270
Lys	Gln	Ser	Ser	Thr	Leu	Thr	Thr	His	Lys	Ile	Ile	His	Thr	Gly
				275					280					285
Glu	Lys	Pro	Tyr	Arg	Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe	Asn	Arg
				290					295					300
Ser	Ser	His	Leu	Thr	Thr	His	Lys	Ile	Ile	His	Thr	Gly	Glu	Lys
				305					310					315
Pro	Tyr	Lys	Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe	Asn	Gln	Ser	Ser
				320					325					330
Thr	Leu	Ser	Thr	His	Lys	Phe	Ile	His	Ala	Gly	Glu	Lys	Pro	Tyr
				335					340					345
Lys	Cys	Glu	Glu	Cys	Asp	Lys	Ala	Phe	Asn	Arg	Phe	Ser	Tyr	Leu
				350					355					360
Thr	Lys	His	Lys	Ile	Ile	His	Thr	Gly	Glu	Lys	Ser	Tyr	Lys	Cys
				365					370					375
Glu	Glu	Cys	Gly	Lys	Gly	Phe	Asn	Trp	Ser	Ser	Thr	Leu	Thr	Lys
				380					385					390
His	Lys	Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Glu	Val
				395					400					405
Cys	Gly	Lys	Ala	Phe	Asn	Glu	Ser	Ser	Asn	Leu	Thr	Thr	His	Lys
				410					415					420
Met	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Glu	Glu	Cys	Gly
				425					430					435
Lys	Ala	Phe	Asn	Arg	Ser	Pro	Gln	Leu	Thr	Ala	His	Lys	Ile	Ile
				440					445					450
His	Thr	Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Glu	Glu	Cys	Gly	Lys	Ala
				455					460					465
Phe	Ser	Gln	Ser	Ser	Ile	Leu	Thr	Thr	His	Lys	Arg	Ile	His	Thr

				470					475					480
Gly	Glu	Lys	Pro	Tyr	Lys	Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe	Asn
				485					490					495
Arg	Ser	Ser	Asn	Leu	Thr	Lys	His	Lys	Ile	Ile	His	Thr	Gly	Glu
				500					505					510
Lys	Ser	Tyr	Lys	Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe	Asn	Gln	Ser
				515					520					525
Ser	Thr	Leu	Thr	Lys	His	Arg	Lys	Ile	His	Thr	Arg	Gln	Lys	Pro
				530					535					540
Tyr	Asn	Cys	Glu	Glu	Cys	Asp	Asn	Thr	Phe	Asn	Gln	Ser	Ser	Asn
				545					550					555
Leu	Ile	Lys	Gln	Asn	Asn	Ser	Tyr	Trp	Arg	Glu	Thr	Leu	Gln	Met
				560					565					570
Ser	Arg	Met	Trp	Glu	Ser	Leu								
				575										

&lt;210&gt; 4

&lt;211&gt; 510

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70287215CD1

&lt;400&gt; 4

Met	Ala	Arg	Glu	Leu	Ser	Glu	Ser	Thr	Ala	Leu	Asp	Ala	Gln	Ser
1				5					10					15
Thr	Glu	Asp	Gln	Met	Glu	Leu	Leu	Val	Ile	Lys	Val	Glu	Glu	Glu
				20					25					30
Glu	Ala	Gly	Phe	Pro	Ser	Ser	Pro	Asp	Leu	Gly	Ser	Glu	Gly	Ser
				35					40					45
Arg	Glu	Arg	Phe	Arg	Gly	Phe	Arg	Tyr	Pro	Glu	Ala	Ala	Gly	Pro
				50					55					60
Arg	Glu	Ala	Leu	Ser	Arg	Leu	Arg	Glu	Leu	Cys	Arg	Gln	Trp	Leu
				65					70					75
Gln	Pro	Glu	Met	His	Ser	Lys	Glu	Gln	Ile	Leu	Glu	Leu	Leu	Val
				80					85					90
Leu	Glu	Gln	Phe	Leu	Thr	Ile	Leu	Pro	Gly	Asn	Leu	Gln	Ser	Trp
				95					100					105
Val	Arg	Glu	Gln	His	Pro	Glu	Ser	Gly	Glu	Glu	Val	Val	Val	Leu
				110					115					120
Leu	Glu	Tyr	Leu	Glu	Arg	Gln	Leu	Asp	Glu	Pro	Ala	Pro	Gln	Val
				125					130					135
Glu	Arg	Thr	Gly	Leu	Val	Ser	Glu	Arg	Cys	Gly	Leu	Phe	Pro	Pro
				140					145					150
Glu	Ile	Pro	Asp	Gln	Ile	Leu	Gln	Val	Pro	Val	Leu	Ala	His	Gly
				155					160					165
Gly	Cys	Cys	Arg	Glu	Asp	Lys	Val	Val	Ala	Ser	Arg	Leu	Thr	Pro
				170					175					180
Glu	Ser	Gln	Gly	Leu	Leu	Lys	Val	Glu	Asp	Val	Ala	Leu	Thr	Leu
				185					190					195
Thr	Pro	Glu	Trp	Thr	Gln	Gln	Asp	Ser	Ser	Gln	Gly	Asn	Leu	Cys
				200					205					210
Arg	Asp	Glu	Lys	Gln	Glu	Asn	His	Gly	Ser	Leu	Val	Phe	Leu	Gly
				215					220					225
Asp	Glu	Lys	Gln	Thr	Lys	Ser	Arg	Asp	Leu	Pro	Pro	Ala	Glu	Glu
				230					235					240

Leu	Pro	Glu	Lys	Glu	His	Gly	Lys	Ile	Ser	Cys	His	Leu	Arg	Glu
				245					250					255
Asp	Ile	Ala	Gln	Ile	Pro	Thr	Cys	Ala	Glu	Ala	Gly	Glu	Gln	Glu
				260					265					270
Gly	Arg	Leu	Gln	Arg	Lys	Gln	Lys	Asn	Ala	Thr	Gly	Gly	Arg	Arg
				275					280					285
His	Ile	Cys	His	Glu	Cys	Gly	Lys	Ser	Phe	Ala	Gln	Ser	Ser	Gly
				290					295					300
Leu	Ser	Lys	His	Arg	Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu
				305					310					315
Cys	Glu	Glu	Cys	Gly	Lys	Ala	Phe	Ile	Gly	Ser	Ser	Ala	Leu	Val
				320					325					330
Ile	His	Gln	Arg	Val	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Glu
				335					340					345
Glu	Cys	Gly	Lys	Ala	Phe	Ser	His	Ser	Ser	Asp	Leu	Ile	Lys	His
				350					355					360
Gln	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Asp	Asp	Cys
				365					370					375
Gly	Lys	Thr	Phe	Ser	Gln	Ser	Cys	Ser	Leu	Leu	Glu	His	His	Arg
				380					385					390
Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Gln	Cys	Ser	Met	Cys	Gly	Lys
				395					400					405
Ala	Phe	Arg	Arg	Ser	Ser	His	Leu	Leu	Arg	His	Gln	Arg	Ile	His
				410					415					420
Thr	Gly	Asp	Lys	Asn	Val	Gln	Glu	Pro	Glu	Gln	Gly	Glu	Ala	Trp
				425					430					435
Lys	Ser	Arg	Met	Glu	Ser	Gln	Leu	Glu	Asn	Val	Glu	Thr	Pro	Met
				440					445					450
Ser	Tyr	Lys	Cys	Asn	Glu	Cys	Glu	Arg	Ser	Phe	Thr	Gln	Asn	Thr
				455					460					465
Gly	Leu	Ile	Glu	His	Gln	Lys	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr
				470					475					480
Gln	Cys	Asn	Ala	Cys	Gly	Lys	Gly	Phe	Thr	Arg	Ile	Ser	Tyr	Leu
				485					490					495
Val	Gln	His	Gln	Arg	Ser	His	Val	Gly	Lys	Asn	Ile	Leu	Ser	Gln
				500					505					510

&lt;210&gt; 5

&lt;211&gt; 448

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 5500054CD1

&lt;400&gt; 5

Met	Ser	Leu	Val	Asp	Leu	Gly	Lys	Arg	Leu	Leu	Glu	Ala	Ala	Arg
1				5					10					15
Lys	Gly	Gln	Asp	Asp	Glu	Val	Arg	Thr	Leu	Met	Ala	Asn	Gly	Ala
				20					25					30
Pro	Phe	Thr	Thr	Asp	Trp	Leu	Gly	Thr	Ser	Pro	Leu	His	Leu	Ala
				35					40					45
Ala	Gln	Tyr	Gly	His	Tyr	Ser	Thr	Ala	Glu	Val	Leu	Leu	Arg	Ala
				50					55					60
Gly	Val	Ser	Arg	Asp	Ala	Arg	Thr	Lys	Val	Asp	Arg	Thr	Pro	Leu
				65					70					75

His	Met	Ala	Ala	Ala	Asp	Gly	His	Ala	His	Ile	Val	Glu	Leu	Leu	80	85	90
Val	Arg	Asn	Gly	Ala	Asp	Val	Asn	Ala	Lys	Asp	Met	Leu	Lys	Met	95	100	105
Thr	Ala	Leu	His	Trp	Ala	Thr	Glu	Arg	His	His	Arg	Asp	Val	Val	110	115	120
Glu	Leu	Leu	Ile	Lys	Tyr	Gly	Ala	Asp	Val	His	Ala	Phe	Ser	Lys	125	130	135
Phe	Asp	Lys	Ser	Ala	Phe	Asp	Ile	Ala	Leu	Glu	Lys	Asn	Asn	Ala	140	145	150
Glu	Ile	Leu	Val	Ile	Leu	Gln	Glu	Ala	Met	Gln	Asn	Gln	Val	Asn	155	160	165
Val	Asn	Pro	Glu	Arg	Ala	Asn	Pro	Val	Thr	Asp	Pro	Val	Ser	Met	170	175	180
Ala	Ala	Pro	Phe	Ile	Phe	Thr	Ser	Gly	Glu	Val	Val	Asn	Leu	Ala	185	190	195
Ser	Leu	Ile	Ser	Ser	Thr	Asn	Thr	Lys	Thr	Thr	Ser	Gly	Asp	Pro	200	205	210
His	Ala	Ser	Thr	Val	Gln	Phe	Ser	Asn	Ser	Thr	Thr	Ser	Val	Leu	215	220	225
Ala	Thr	Leu	Ala	Ala	Leu	Ala	Glu	Ala	Ser	Val	Pro	Leu	Ser	Asn	230	235	240
Ser	His	Arg	Ala	Thr	Ala	Asn	Thr	Glu	Glu	Ile	Ile	Glu	Gly	Asn	245	250	255
Ser	Val	Asp	Ser	Ser	Ile	Gln	Gln	Val	Met	Gly	Ser	Gly	Gly	Gln	260	265	270
Arg	Val	Ile	Thr	Ile	Val	Thr	Asp	Gly	Val	Pro	Leu	Gly	Asn	Ile	275	280	285
Gln	Thr	Ser	Ile	Pro	Thr	Gly	Gly	Ile	Gly	Gln	Pro	Phe	Ile	Val	290	295	300
Thr	Val	Gln	Asp	Gly	Gln	Gln	Val	Leu	Thr	Val	Pro	Ala	Gly	Lys	305	310	315
Val	Ala	Glu	Glu	Thr	Val	Ile	Lys	Glu	Glu	Glu	Glu	Glu	Lys	Leu	320	325	330
Pro	Leu	Thr	Lys	Lys	Pro	Arg	Ile	Gly	Glu	Lys	Thr	Asn	Ser	Val	335	340	345
Glu	Glu	Ser	Lys	Glu	Gly	Asn	Glu	Arg	Glu	Leu	Leu	Gln	Gln	Gln	350	355	360
Leu	Gln	Glu	Ala	Asn	Arg	Arg	Ala	Gln	Glu	Tyr	Arg	His	Gln	Leu	365	370	375
Leu	Lys	Lys	Glu	Gln	Glu	Ala	Glu	Gln	Tyr	Arg	Leu	Lys	Leu	Glu	380	385	390
Ala	Ile	Ala	Arg	Gln	Gln	Pro	Asn	Gly	Val	Asp	Phe	Thr	Met	Val	395	400	405
Glu	Glu	Val	Ala	Glu	Val	Asp	Ala	Val	Val	Val	Thr	Glu	Gly	Glu	410	415	420
Leu	Glu	Glu	Arg	Glu	Thr	Lys	Val	Thr	Gly	Ser	Ala	Gly	Thr	Thr	425	430	435
Glu	Pro	His	Thr	Arg	Val	Ser	Met	Ala	Thr	Val	Ser	Ser			440	445	

&lt;210&gt; 6

&lt;211&gt; 247

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 1349648CD1

&lt;400&gt; 6

Met	Ala	Asp	Thr	Thr	Pro	Asn	Gly	Pro	Gln	Gly	Ala	Gly	Ala	Val
1				5					10					15
Gln	Phe	Met	Met	Thr	Asn	Lys	Leu	Asp	Thr	Ala	Met	Trp	Leu	Ser
				20					25					30
Arg	Leu	Phe	Thr	Val	Tyr	Cys	Ser	Ala	Leu	Phe	Val	Leu	Pro	Leu
				35					40					45
Leu	Gly	Leu	His	Glu	Ala	Ala	Ser	Phe	Tyr	Gln	Arg	Ala	Leu	Leu
				50					55					60
Ala	Asn	Ala	Leu	Thr	Ser	Ala	Leu	Arg	Leu	His	Gln	Arg	Leu	Pro
				65					70					75
His	Phe	Gln	Leu	Ser	Arg	Ala	Phe	Leu	Ala	Gln	Ala	Leu	Leu	Glu
				80					85					90
Asp	Ser	Cys	His	Tyr	Leu	Leu	Tyr	Ser	Leu	Ile	Phe	Val	Asn	Ser
				95					100					105
Tyr	Pro	Val	Thr	Met	Ser	Ile	Phe	Pro	Val	Leu	Leu	Phe	Ser	Leu
				110					115					120
Leu	His	Ala	Ala	Thr	Tyr	Thr	Lys	Lys	Val	Leu	Asp	Ala	Arg	Gly
				125					130					135
Ser	Asn	Ser	Leu	Pro	Leu	Leu	Arg	Ser	Val	Leu	Asp	Lys	Leu	Ser
				140					145					150
Ala	Asn	Gln	Gln	Asn	Ile	Leu	Lys	Phe	Ile	Ala	Cys	Asn	Glu	Ile
				155					160					165
Phe	Leu	Met	Pro	Ala	Thr	Val	Phe	Met	Leu	Phe	Ser	Gly	Gln	Gly
				170					175					180
Ser	Leu	Leu	Gln	Pro	Phe	Ile	Tyr	Tyr	Arg	Phe	Leu	Thr	Leu	Arg
				185					190					195
Tyr	Ser	Ser	Arg	Arg	Asn	Pro	Tyr	Cys	Arg	Thr	Leu	Phe	Asn	Glu
				200					205					210
Leu	Arg	Ile	Val	Val	Glu	His	Ile	Ile	Met	Lys	Pro	Ala	Cys	Pro
				215					220					225
Leu	Phe	Val	Arg	Arg	Leu	Cys	Leu	Gln	Ser	Ile	Ala	Phe	Ile	Ser
				230					235					240
Arg	Leu	Ala	Pro	Thr	Val	Pro								
				245										

&lt;210&gt; 7

&lt;211&gt; 636

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3888314CD1

&lt;400&gt; 7

Met	Ala	Pro	Arg	Pro	Pro	Thr	Ala	Ala	Pro	Gln	Glu	Ser	Val	Thr
1				5					10					15
Phe	Lys	Asp	Val	Ser	Val	Asp	Phe	Thr	Gln	Glu	Glu	Trp	Tyr	His
				20					25					30
Val	Asp	Pro	Ala	Gln	Arg	Ser	Leu	Tyr	Arg	Asp	Val	Met	Leu	Glu
				35					40					45
Asn	Tyr	Ser	His	Leu	Val	Ser	Leu	Gly	Tyr	Gln	Val	Ser	Lys	Pro
				50					55					60
Glu	Val	Ile	Phe	Lys	Leu	Glu	Gln	Gly	Glu	Glu	Pro	Trp	Ile	Ser
				65					70					75

Glu	Gly	Glu	Ile	Gln	Arg	Pro	Phe	Tyr	Pro	Asp	Trp	Lys	Thr	Arg	80	85	90
Pro	Glu	Val	Lys	Ser	Ser	His	Leu	Gln	Gln	Asp	Val	Ser	Glu	Val	95	100	105
Ser	His	Cys	Thr	His	Asp	Leu	Leu	His	Ala	Thr	Leu	Glu	Asp	Ser	110	115	120
Trp	Asp	Val	Ser	Ser	Gln	Leu	Asp	Arg	Gln	Gln	Glu	Asn	Trp	Lys	125	130	135
Arg	His	Leu	Gly	Ser	Glu	Ala	Ser	Thr	Gln	Lys	Lys	Ile	Ile	Thr	140	145	150
Pro	Gln	Glu	Asn	Phe	Glu	Gln	Asn	Lys	Phe	Gly	Glu	Asn	Ser	Arg	155	160	165
Leu	Asn	Thr	Asn	Leu	Val	Thr	Gln	Leu	Asn	Ile	Pro	Ala	Arg	Ile	170	175	180
Arg	Pro	Ser	Glu	Cys	Glu	Thr	Leu	Gly	Ser	Asn	Leu	Gly	His	Asn	185	190	195
Ala	Asp	Leu	Leu	Asn	Glu	Asn	Asn	Ile	Leu	Ala	Lys	Lys	Lys	Pro	200	205	210
Tyr	Lys	Cys	Asp	Lys	Cys	Arg	Lys	Ala	Phe	Ile	His	Arg	Ser	Ser	215	220	225
Leu	Thr	Lys	His	Glu	Lys	Thr	His	Lys	Gly	Glu	Gly	Ala	Phe	Pro	230	235	240
Asn	Gly	Thr	Asp	Gln	Gly	Ile	Tyr	Pro	Gly	Lys	Lys	His	His	Glu	245	250	255
Cys	Thr	Asp	Cys	Gly	Lys	Thr	Phe	Leu	Trp	Lys	Thr	Gln	Leu	Thr	260	265	270
Glu	His	Gln	Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Phe	Glu	Cys	Asn	275	280	285
Val	Cys	Gly	Lys	Ala	Phe	Arg	His	Ser	Ser	Ser	Leu	Gly	Gln	His	290	295	300
Glu	Asn	Ala	His	Thr	Gly	Glu	Lys	Pro	Tyr	Gln	Cys	Ser	Leu	Cys	305	310	315
Gly	Lys	Ala	Phe	Gln	Arg	Ser	Ser	Ser	Leu	Val	Gln	His	Gln	Arg	320	325	330
Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Arg	Cys	Asn	Leu	Cys	Gly	Arg	335	340	345
Ser	Phe	Arg	His	Gly	Thr	Ser	Leu	Thr	Gln	His	Glu	Val	Thr	His	350	355	360
Ser	Gly	Glu	Lys	Pro	Phe	Gln	Cys	Lys	Glu	Cys	Gly	Lys	Ala	Phe	365	370	375
Ser	Arg	Cys	Ser	Ser	Leu	Val	Gln	His	Glu	Arg	Thr	His	Thr	Gly	380	385	390
Glu	Lys	Pro	Phe	Glu	Cys	Ser	Ile	Cys	Gly	Arg	Ala	Phe	Gly	Gln	395	400	405
Ser	Pro	Ser	Leu	Tyr	Lys	His	Met	Arg	Ile	His	Lys	Arg	Gly	Lys	410	415	420
Pro	Tyr	Gln	Ser	Ser	Asn	Tyr	Ser	Ile	Asp	Phe	Lys	His	Ser	Thr	425	430	435
Ser	Leu	Thr	Gln	Asp	Glu	Ser	Thr	Leu	Thr	Glu	Val	Lys	Ser	Tyr	440	445	450
His	Cys	Asn	Asp	Cys	Gly	Glu	Asp	Phe	Ser	His	Ile	Thr	Asp	Phe	455	460	465
Thr	Asp	His	Gln	Arg	Ile	His	Thr	Ala	Glu	Asn	Pro	Tyr	Asp	Cys	470	475	480
Glu	Gln	Ala	Phe	Ser	Gln	Gln	Ala	Ile	Ser	His	Pro	Gly	Glu	Lys	485	490	495
Pro	Tyr	Gln	Cys	Asn	Val	Cys	Gly	Lys	Ala	Phe	Lys	Arg	Ser	Thr	500	505	510



Ser	Phe	Ile	Glu	His	His	Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr
				515					520					525
Glu	Cys	Asn	Glu	Cys	Gly	Glu	Ala	Phe	Ser	Arg	Arg	Ser	Ser	Leu
				530					535					540
Thr	Gln	His	Glu	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys
				545					550					555
Ile	Asp	Cys	Gly	Lys	Ala	Phe	Ser	Gln	Ser	Ser	Ser	Leu	Ile	Gln
				560					565					570
His	Glu	Arg	Thr	His	Thr	Gly	Glu	Lys	Pro	Tyr	Glu	Cys	Asn	Glu
				575					580					585
Cys	Gly	Arg	Ala	Phe	Arg	Lys	Lys	Thr	Asn	Leu	His	Asp	His	Gln
				590					595					600
Arg	Ile	His	Thr	Gly	Glu	Lys	Pro	Tyr	Ser	Cys	Lys	Glu	Cys	Gly
				605					610					615
Lys	Asn	Phe	Ser	Arg	Ser	Ser	Ala	Leu	Thr	Lys	His	Gln	Arg	Ile
				620					625					630
His	Thr	Arg	Asn	Lys	Leu									
				635										

&lt;210&gt; 8

&lt;211&gt; 1603

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3276670CD1

&lt;400&gt; 8

Met	Arg	Val	Lys	Pro	Gln	Gly	Leu	Val	Val	Thr	Ser	Ser	Ala	Val
1				5					10					15
Cys	Ser	Ser	Pro	Asp	Tyr	Leu	Arg	Glu	Pro	Lys	Tyr	Tyr	Pro	Gly
				20					25					30
Gly	Pro	Pro	Thr	Pro	Arg	Pro	Leu	Leu	Pro	Thr	Arg	Pro	Pro	Ala
				35					40					45
Ser	Pro	Pro	Asp	Lys	Ala	Phe	Ser	Thr	His	Ala	Phe	Ser	Glu	Asn
				50					55					60
Pro	Arg	Pro	Pro	Pro	Arg	Arg	Asp	Pro	Ser	Thr	Arg	Arg	Pro	Pro
				65					70					75
Val	Leu	Ala	Lys	Gly	Asp	Asp	Pro	Leu	Pro	Pro	Arg	Ala	Ala	Arg
				80					85					90
Pro	Val	Ser	Gln	Ala	Arg	Cys	Pro	Thr	Pro	Val	Gly	Asp	Gly	Ser
				95					100					105
Ser	Ser	Arg	Arg	Cys	Trp	Asp	Asn	Gly	Arg	Val	Asn	Leu	Arg	Pro
				110					115					120
Val	Val	Gln	Leu	Ile	Asp	Ile	Met	Lys	Asp	Leu	Thr	Arg	Leu	Ser
				125					130					135
Gln	Asp	Leu	Gln	His	Ser	Gly	Val	His	Leu	Asp	Cys	Gly	Gly	Leu
				140					145					150
Arg	Leu	Ser	Arg	Pro	Pro	Ala	Pro	Pro	Pro	Gly	Asp	Leu	Gln	Tyr
				155					160					165
Ser	Phe	Phe	Ser	Ser	Pro	Ser	Leu	Ala	Asn	Ser	Ile	Arg	Ser	Pro
				170					175					180
Glu	Glu	Arg	Ala	Thr	Pro	His	Ala	Lys	Ser	Glu	Arg	Pro	Ser	His
				185					190					195
Pro	Leu	Tyr	Glu	Pro	Glu	Pro	Glu	Pro	Arg	Asp	Ser	Pro	Gln	Pro
				200					205					210
Gly	Gln	Gly	His	Ser	Pro	Gly	Ala	Thr	Ala	Ala	Ala	Thr	Gly	Leu

	215		220		225
Pro Pro Glu Pro	Glu Pro Asp Ser Thr	Asp Tyr Ser Glu Leu	Ala		
	230		235		240
Asp Ala Asp Ile	Leu Ser Glu Leu Ala	Ser Leu Thr Cys Pro	Glu		
	245		250		255
Ala Gln Leu Leu	Glu Ala Gln Ala Leu	Glu Pro Pro Ser Pro	Glu		
	260		265		270
Pro Glu Pro Gln	Leu Leu Asp Pro Gln	Pro Arg Phe Leu Asp	Pro		
	275		280		285
Gln Ala Leu Glu	Pro Leu Gly Glu Ala	Leu Glu Leu Pro Pro	Leu		
	290		295		300
Gln Pro Leu Ala	Asp Pro Leu Gly Leu	Pro Gly Leu Ala Leu	Gln		
	305		310		315
Ala Leu Asp Thr	Leu Pro Asp Ser Leu	Glu Ser Gln Leu Leu	Asp		
	320		325		330
Pro Gln Ala Leu	Asp Pro Leu Pro Lys	Leu Leu Asp Val Pro	Gly		
	335		340		345
Arg Arg Leu Glu	Pro Gln Gln Pro Leu	Gly His Cys Pro Leu	Ala		
	350		355		360
Glu Pro Leu Arg	Leu Asp Leu Cys Ser	Pro His Gly Pro Pro	Gly		
	365		370		375
Pro Glu Gly His	Pro Lys Tyr Ala Leu	Arg Arg Thr Asp Arg	Pro		
	380		385		390
Lys Ile Leu Cys	Arg Arg Arg Lys Ala	Gly Arg Gly Arg Lys	Ala		
	395		400		405
Asp Ala Gly Pro	Glu Gly Arg Leu Leu	Pro Leu Pro Met Pro	Thr		
	410		415		420
Gly Leu Val Ala	Ala Leu Ala Glu Pro	Pro Pro Pro Pro Pro	Pro		
	425		430		435
Pro Pro Pro Ala	Leu Pro Gly Pro Gly	Pro Val Ser Val Pro	Glu		
	440		445		450
Leu Lys Pro Glu	Ser Ser Gln Thr Pro	Val Val Ser Thr Arg	Lys		
	455		460		465
Gly Lys Cys Arg	Gly Val Arg Arg Met	Val Val Lys Met Ala	Lys		
	470		475		480
Ile Pro Val Ser	Leu Gly Arg Arg Asn	Lys Thr Thr Tyr Lys	Val		
	485		490		495
Ser Ser Leu Ser	Ser Ser Leu Ser Val	Glu Gly Lys Glu Leu	Gly		
	500		505		510
Leu Arg Val Ser	Ala Glu Pro Thr Pro	Leu Leu Lys Met Lys	Asn		
	515		520		525
Asn Gly Arg Asn	Val Val Val Val Phe	Pro Pro Gly Glu Met	Pro		
	530		535		540
Ile Ile Leu Lys	Arg Lys Arg Gly Arg	Pro Pro Lys Asn Leu	Leu		
	545		550		555
Leu Gly Pro Gly	Lys Pro Lys Glu Pro	Ala Val Val Ala Ala	Glu		
	560		565		570
Ala Ala Thr Val	Ala Ala Ala Thr Met	Ala Met Pro Glu Val	Lys		
	575		580		585
Lys Arg Arg Arg	Arg Lys Gln Lys Leu	Ala Ser Pro Gln Pro	Ser		
	590		595		600
Tyr Ala Ala Asp	Ala Asn Asp Ser Lys	Ala Glu Tyr Ser Asp	Val		
	605		610		615
Leu Ala Lys Leu	Ala Phe Leu Asn Arg	Gln Ser Gln Cys Ala	Gly		
	620		625		630
Arg Cys Ser Pro	Pro Arg Cys Trp Thr	Pro Ser Glu Pro Glu	Ser		
	635		640		645
Val His Gln Ala	Pro Asp Thr Gln Ser	Ile Ser His Phe Leu	His		

Arg Val Gln Gly	650	Arg Arg Arg Gly	655	Lys Ala Gly Gly	660
Phe	665	Gly	670	Phe	675
Gly Gly Arg Gly	680	Gly Gly His Ala Ala	685	Lys Ser Ala Arg Cys Ser	690
Phe Ser Asp Phe	695	Phe Glu Gly Ile Gly	700	Lys Lys Lys Lys Val Val	705
Ala Val Ala Ala	710	Ala Gly Val Gly Gly	715	Pro Gly Leu Thr Glu Leu	720
Gly His Pro Arg	725	Lys Arg Gly Arg Gly	730	Glu Val Asp Ala Val Thr	735
Gly Lys Pro Lys	740	Arg Lys Arg Arg Ser	745	Arg Lys Asn Gly Thr Leu	750
Phe Pro Glu Gln	755	Val Pro Ser Gly Pro	760	Gly Phe Gly Glu Ala Gly	765
Ala Glu Trp Ala	770	Gly Asp Lys Gly Gly	775	Gly Trp Ala Pro His His	780
Gly His Pro Gly	785	Gly Gln Ala Gly Arg	790	Asn Cys Gly Phe Gln Gly	795
Thr Glu Ala Arg	800	Ala Phe Ala Ser Thr	805	Gly Leu Glu Ser Gly Ala	810
Ser Gly Arg Gly	815	Ser Tyr Tyr Ser Thr	820	Gly Ala Pro Ser Gly Gln	825
Thr Glu Leu Ser	830	Gln Glu Arg Gln Asn	835	Leu Phe Thr Gly Tyr Phe	840
Arg Ser Leu Leu	845	Asp Ser Asp Asp Ser	850	Ser Asp Leu Leu Asp Phe	855
Ala Leu Ser Ala	860	Ser Arg Pro Glu Ser	865	Arg Lys Ala Ser Gly Thr	870
Tyr Ala Gly Pro	875	Pro Thr Ser Ala Leu	880	Pro Ala Gln Arg Gly Leu	885
Ala Thr Phe Pro	890	Ser Arg Gly Ala Lys	895	Ala Ser Pro Val Ala Val	900
Gly Ser Ser Gly	905	Ala Gly Ala Asp Pro	910	Ser Phe Gln Pro Val Leu	915
Ser Ala Arg Gln	920	Thr Phe Pro Pro Gly	925	Arg Ala Ala Ser Tyr Gly	930
Leu Thr Pro Ala	935	Thr Ser Asp Cys Arg	940	Ala Ala Glu Thr Phe Pro	945
Lys Leu Val Pro	950	Pro Pro Ser Ala Met	955	Ala Arg Ser Pro Thr Thr	960
His Pro Pro Ala	965	Asn Thr Tyr Leu Pro	970	Gln Tyr Gly Gly Tyr Gly	975
Ala Gly Gln Ser	980	Val Phe Ala Pro Thr	985	Lys Pro Phe Thr Gly Gln	990
Asp Cys Ala Asn	995	Ser Lys Asp Cys Ser	1000	Phe Ala Tyr Gly Ser Gly	1005
Asn Ser Leu Pro	1010	Ala Ser Pro Ser Ser	1015	Ala His Ser Ala Gly Tyr	1020
Ala Pro Pro Pro	1025	Thr Gly Gly Pro Cys	1030	Leu Pro Pro Ser Lys Ala	1035
Ser Phe Phe Ser	1040	Ser Ser Glu Gly Ala	1045	Pro Phe Ser Gly Ser Ala	1050
Pro Thr Pro Leu	1055	Arg Cys Asp Ser Arg	1060	Ala Ser Thr Val Ser Pro	1065
Gly Gly Tyr Met	1070	Val Pro Lys Gly Thr	1075	Thr Ala Ser Ala Thr Ser	1080
Ala Ala Ser Ala		Ala Ser Ser Ser Ser		Ser Ser Phe Gln Pro Ser	

1085	1090	1095
Pro Glu Asn Cys Arg Gln Phe Ala Gly Ala Ser Gln Trp Pro Phe		
1100	1105	1110
Arg Gln Gly Tyr Gly Gly Leu Asp Trp Ala Ser Glu Ala Phe Ser		
1115	1120	1125
Gln Leu Tyr Asn Pro Ser Phe Asp Cys His Val Ser Glu Pro Asn		
1130	1135	1140
Val Ile Leu Asp Ile Ser Asn Tyr Thr Pro Gln Lys Val Lys Gln		
1145	1150	1155
Gln Thr Ala Val Ser Glu Thr Phe Ser Glu Ser Ser Ser Asp Ser		
1160	1165	1170
Thr Gln Phe Asn Gln Pro Val Gly Gly Gly Gly Phe Arg Arg Ala		
1175	1180	1185
Asn Ser Glu Ala Ser Ser Ser Glu Gly Gln Ser Ser Leu Ser Ser		
1190	1195	1200
Leu Glu Lys Leu Met Met Asp Trp Asn Glu Ala Ser Ser Ala Pro		
1205	1210	1215
Gly Tyr Asn Trp Asn Gln Ser Val Leu Phe Gln Ser Ser Ser Lys		
1220	1225	1230
Pro Gly Arg Gly Arg Arg Lys Lys Val Asp Leu Phe Glu Ala Ser		
1235	1240	1245
His Leu Gly Phe Pro Thr Ser Ala Ser Ala Ala Ala Ser Gly Tyr		
1250	1255	1260
Pro Ser Lys Arg Ser Thr Gly Pro Arg Gln Pro Arg Gly Gly Arg		
1265	1270	1275
Gly Gly Gly Ala Cys Ser Ala Lys Lys Glu Arg Gly Gly Ala Ala		
1280	1285	1290
Ala Lys Ala Lys Phe Ile Pro Lys Pro Gln Pro Val Asn Pro Leu		
1295	1300	1305
Phe Gln Asp Ser Pro Asp Leu Gly Leu Asp Tyr Tyr Ser Gly Asp		
1310	1315	1320
Ser Ser Met Ser Pro Leu Pro Ser Gln Ser Arg Ala Phe Gly Val		
1325	1330	1335
Gly Glu Arg Asp Pro Cys Asp Phe Ile Gly Pro Tyr Ser Met Asn		
1340	1345	1350
Pro Ser Thr Pro Ser Asp Gly Thr Phe Gly Gln Gly Phe His Cys		
1355	1360	1365
Asp Ser Pro Ser Leu Gly Ala Pro Glu Leu Asp Gly Lys His Phe		
1370	1375	1380
Pro Pro Leu Ala His Pro Pro Thr Val Phe Asp Ala Gly Leu Gln		
1385	1390	1395
Lys Ala Tyr Ser Pro Thr Cys Ser Pro Thr Leu Gly Phe Lys Glu		
1400	1405	1410
Glu Leu Arg Pro Pro Pro Thr Lys Leu Ala Ala Cys Glu Pro Leu		
1415	1420	1425
Lys His Gly Leu Gln Gly Ala Ser Leu Gly His Ala Ala Ala Ala		
1430	1435	1440
Gln Ala His Leu Ser Cys Arg Asp Leu Pro Leu Gly Gln Pro His		
1445	1450	1455
Tyr Asp Ser Pro Ser Cys Lys Gly Thr Ala Tyr Trp Tyr Pro Pro		
1460	1465	1470
Gly Ser Ala Ala Arg Ser Pro Pro Tyr Glu Gly Lys Val Gly Thr		
1475	1480	1485
Gly Leu Leu Ala Asp Phe Leu Gly Arg Thr Glu Ala Ala Cys Leu		
1490	1495	1500
Ser Ala Pro His Leu Ala Ser Pro Pro Ala Thr Pro Lys Ala Asp		
1505	1510	1515
Lys Glu Pro Leu Glu Met Ala Arg Pro Pro Gly Pro Pro Arg Gly		

1520	1525	1530
Pro Ala Ala Ala Ala Ala Gly Tyr Gly Cys Pro Leu Leu Ser Asp		
1535	1540	1545
Leu Thr Leu Ser Pro Val Pro Arg Asp Ser Leu Leu Pro Leu Gln		
1550	1555	1560
Asp Thr Ala Tyr Arg Tyr Pro Gly Phe Met Pro Gln Ala His Pro		
1565	1570	1575
Gly Leu Gly Gly Gly Pro Lys Ser Gly Phe Leu Gly Pro Met Ala		
1580	1585	1590
Glu Pro His Pro Glu Asp Thr Phe Thr Val Thr Ser Leu		
1595	1600	

<210> 9  
 <211> 1629  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2729623CB1

<400> 9

ggcgcctcgg	actttttgctc	ccacaagtcc	tgccctcggag	gcggggggagc	tggaaccagca	60
gccgcctgga	gcgtccgagt	caccgtcggc	ggggctcccc	cgctccccag	aacgggtggga	120
cgcgggggctc	ggcagccgcc	agcgggaacat	ggcgccctgg	acgctgtggc	gctgctgcca	180
gcgcgtcgtg	ggctgggtgc	cggtgctctt	catcaccttc	gtggctcgtct	ggtcctacta	240
cgcgtagctg	gtggagctct	gcgtgtttac	tattttttgga	aatgaagaaa	atggaaagac	300
cgttggtttac	cttgtgggctt	tccatctggt	ctttgttatg	tttgtatggg	cctattggat	360
gacaatttttc	acatctcccc	cttccccctc	caaagagttc	tacttgtcca	attctgaaaa	420
ggaacgttat	gaaaaagaat	tcagccaaga	aagacaacaa	gaaattttga	gaagagcagc	480
aagagcttta	cctatctata	ccacatcagc	ttcaaaaact	atcagatatt	gtgaaaaatg	540
tcagctgatt	aaacctgac	gggcgcac	ctgctcagcc	tgtgactcat	gtattcttaa	600
gatggatcat	cactgtcctt	gggtgaataa	ctgtgtggga	ttttctaatt	acaaattctt	660
cctgctgttt	ttattgtatt	ccctattata	ttgccttttc	gtggctgcaa	cagtttttaga	720
gtactttata	aaatttttga	cgaatgaact	gacagataca	cgtgcaaaat	tccacgtact	780
ttttcttttt	tttgtgtctg	caatgttctt	catcagcgtc	ctctcacttt	tcagctacca	840
ctgctggcta	gttggaaaaa	atagaacaac	aatagaatca	ttccgcgcac	ccacgttttc	900
atacggacct	gatggaaatg	gtttctctct	tggtatgcagt	aaaaatttga	gacaagtctt	960
tggtgatgaa	aagaaatatt	ggctacttcc	aatattttca	agcttgggtg	atggttgcag	1020
ttttccaact	cgccttgtgg	ggatggatcc	agaacaagct	tctgttacaa	accagaatga	1080
gtatgccaga	agtagtggct	caaatacaac	ttttcctatc	aaaccactta	gtgaatcaaa	1140
aaaccgcttg	ttggacagtg	aatctcagtg	gctggagaat	ggagctgaag	aaggcatcgt	1200
caaatcaggt	gtatgaaaac	attatagact	ggtattttca	attttcattt	gcaagaaaat	1260
gatcagtggg	atgaaataac	tgaagtataa	cagaagatat	atttttttaa	acggaaagcc	1320
tttgtacagt	tcctgggatt	cacagaagca	ctactccaga	gcagaatgat	gccttaattct	1380
taagtgtcca	tttgtgcagc	attgacttag	agctacaaaa	gtgacttaat	gttattcttg	1440
aaataatact	tacctgttat	gagttgctat	aatatgagct	gtcatcacat	tttaacatgc	1500
atatgtattt	tttgatacct	gaattacatt	attagaataa	gtatccatat	acattttcac	1560
tccaaaaaca	caagcttaaa	aacttttaaag	tacatctagg	gcaaattggtg	gctgaaagtg	1620
aataatatt						1629

<210> 10  
 <211> 3961  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature

&lt;223&gt; Incyte ID No: 7474993CB1

&lt;400&gt; 10

```

cgccgccacg gaccgcctca gccctccggg ccacggggcct ctccggggcgg gagcccaggg 60
tggcgggcgcc tgcggccgag cgccggggagc aggtagtgtt ggtgcttctt ttccagtgcc 120
agtccatcca tcacgaccac atttgtcatg atgacagtga caagatatct cccgggggcca 180
attcagcatc tctacctggc catcctaaca aggtgatttg tgaaaggggtg agacttcaga 240
gcctgttccc tctcctccca agtgatcaga acactaccgt tcaagaggat gctcacttca 300
aagctttctt ccagagtga gtagtccaa gtcctaagag acagcgcctc tctcattcag 360
tctttgatta tacatcagca tcaccagctc cctcaccacc aatgcgacca tgggagatga 420
catcaaatag gcagccccct tcagttcgac caagccaaca tcacttctca ggggaacgat 480
gcaacacacc tgcacgcaac agaagaagtc ctctgtcag gcgccagaga ggaagaaggg 540
atcgtctgtc tcgacataat tccattagtc aagatgaaaa ctatcaccat ctcccttacg 600
cacagcagca agcaatagag gagcctcgag ccttcacacc tccgaatgta tctccccgtc 660
tgctacatcc tgctgctcat ccaccccagc agaatgcagt catggttgac atacatgatc 720
agctccatca aggaacagtc cctgtttctt acacagtaac aacagtggca ccacatggga 780
ttccactctg cacaggccag cacatccctg ctgttagtag acagcagggtc ccaggatgct 840
ctgtgggtttt cagtggacag cacctccctg tctgtagtgt gcctcctcca atgcttcagg 900
catgttcagt tcagcactta ccagtaccat atgctgcatt cccacccctt atttctagt 960
atccatttct tatacatcct cctcaccttt ctccccatca tctcctcat ttgccaccac 1020
caggccagtt tgtccctttc caaacacagc aatcacgata gcctctgcaa aggatagaaa 1080
atgaagtggg actcttagga gaacatcttc cagtaggagg ttttacttac cctccatcag 1140
cccaccccc aacattacct ccacagctc ccttgagtt cttaacacat gatcctttgc 1200
atcaggaggt gtcttttgga gtaccttate ctccatttat gcctcggagg cttacaggac 1260
gtagtagata ccgatcccag cagccaatac cactcctccc ttatcatccc agcttactgc 1320
catatgtgtt atcaatgctt ccagtggcag ctgcagtggg cccaactttc agctttgaat 1380
tagatgtaga agatggagaa gtagaaaatt acgaggccct gttaaacctg gcagagcgac 1440
tgggagaggc aaagcctcgt ggactgacta aagcagatat tgaacaactt ccttcttate 1500
ggttcaatcc taacaaccac cagtcagaac agactttgtg tgtagtatgc atgtgtgatt 1560
ttgagtcaag gcagctactt agagtcttac cctgtaacca cgagttccat gccagtgtg 1620
ttgacaaatg gcttaaggca aatcgtactt gcccaatttg ccgagctgat gcttcagaag 1680
tgcacgggga ttcagaatga ccaacctaa agcacaaaat ttagtgtggg tgttctcat 1740
cacatgtata tacggactat ccattgaact taatctgtgt ggcttcacgc cctcccttta 1800
ccaaaagggt caatggacct ttctttgcac tgtgtgactt aatcaactat aaaagcttac 1860
aattagtctt cacagttagt ggattgttat actaacctgt tgattggaac tccaaagact 1920
ttttcttttag cttaattttg tgtgtgcact aacattccct ggtttttgtg tgatcattcc 1980
gagtgttgct gcaagattac agtggacgtg atcttttagc atgtgctttt ataaaaagtg 2040
gtagctacag atgatatagc aatctacctt atatagagcc ttcagaaact gtgagtggaa 2100
atgaatgccc agcgtatgac tgttggtgat aagtgtatct gtgtgagtgt gtgcaactac 2160
ttgtgtgagg gcatggtagc taacgtgtgt atgagatcct atataccagc atgtaccaag 2220
aatgtgtgtg tagttttaat tatgctgcaa tgtataatct ggtgtgttat ttaaacagca 2280
ctagtactgt acactggttt ttttcccttg tgtttgtctg tgcacactga ttgctgaggg 2340
tgcacatcaatt acaagcattg aatactccat ccccttccct ccagtgatg ggaatgagaa 2400
aagccttctt cctttgcttc aggcagctgt cacttctctt ttattggtgg tcctaagtg 2460
gtcacttaag aaaaaaagt gtaacaaatt ctactccat gaacctggtt tttttttatt 2520
ctgttgtgga aattttttaa tctccaactt gctgtttcca aaaagaagg gcaatatcat 2580
atatcatcag ttagcaatat tagcatttca atcagtgat tgaatgttga tacttttttt 2640
tcctttacat ttcagtaag tttcttacag aaagtacctg tgattttttt taatgttttag 2700
atgtgtagtc tagtctatc tgaagatatt tgattaatat atttctaaga agaccactag 2760
tcccatgaag tctcaccttg ttgacttcca ccagaacaaa acctgttaat tcagcttgac 2820
ttctcatatg tgttctgtgt tcaagctcct gtgggcaact caattgtttt aagaaaactc 2880
tatttctgtt catattaatt cacttgccag actccaattc tgaaagtgtg ctagttcttc 2940
cttcatcaca cgtgcttctc atcgaaacct tggtttcagg atggatcaag tgctgttaca 3000
gctcatgttc cttagacctg atttgttttg atattttact gcctttcctt ttaatttttg 3060
tttgaatgag aaaagtttaa tacaacagg gtccagaata cttgcagtat aaatgaagat 3120
ttggtttttg tataaaaaaa taatgctgtt aaacaggagt tgaccttcat ttagttgatg 3180
taatacatag ctgtatgggg ttttttttta gtaagttctt tgagcataat tcactttaaa 3240
gtcatttttc cagcaatggt taaattactt tctcattctt ttagtgtatt caacattgtc 3300

```

```

tgctcttcc tgcagttgat gtaattgctt tgtttgcaat agcacaagct gcattattcc 3360
agtcaggact gtgataactt gctgccagcc ccactcaact ttcagttggc tctgtgtcag 3420
ttttccactc agtgtttaact acttgttact gccatgctgc ttgccctccc ttgaagtgtc 3480
tataagctca tcacagccta gagttaagta aagtcaattc acagaagcac aattttgccc 3540
tttgcgagac attgttgcct ctatctagtc ctacaagtag ggttttgcat actgtgtttg 3600
cccctagggg tgtcagtgca tcagaaatac ttctaaatag tggtaaaaaat gcacatgggt 3660
aatgcacatg ttacttttaa atcattagga tatccctcac ctgttcctga tgaataaaaa 3720
gtgtgttaaa gaccaaattt cttggcataa taatcagcta catacaaatt acatatagtt 3780
taatcttttt taatgaaaaa aaaatcatgt ttaaaatggc aaaagcccat cttatacact 3840
tttatatagc tgcaaaaaat ttatatctgt acagatctaa cactacgaca ctcagtattc 3900
attttattga agcatgcaag taaagcactt tttctaattt atatagaggt atctaattaa 3960
c

```

&lt;210&gt; 11

&lt;211&gt; 2509

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 55022684CB1

&lt;400&gt; 11

```

gctccagggt tcccccttgc tgctctgtgt cctctgtctc tagaggccca acatctgtgg 60
ccctgtgacc tgcaggtatt gggagaccca cagctaagac accgggaccc cctgaaagcc 120
tagaaatgga cgacttgaaa tatggagtgt atcctctcaa ggaagcaagt ggatgccctg 180
gggctgagag gaatcttcta gtttactctt attttgaaaa ggagacattg acatttaggg 240
atgtggccat agaattctct ctggaggagt gggaatgcct gaaccctgct cagcagaatt 300
tatatatgaa tgtgatgtta gaaaactaca aaaacctggg cttcttggca ggtgttgctg 360
tctctaagca agaccagtc acctgtctgg agcaagaaaa agagccctgg aatatgaaga 420
gacatgagat ggtggatgaa cccccagcta tgtgttctta ttttaccaaa gaccttggc 480
cagagcaaga cataaaagat tcttttcaac aagtaatact gagaagatat ggcaaagtgt 540
aacatgagaa tttacagtta agaaaaggct ccgcaagtgt agatgagtat aagggtgcaca 600
aagaagggtta taatgagcta aaccagtgtt tgacaactac ccagagcaaa atatttccat 660
gtgataaata tgtgaaagtc tttcataaat ttttaaatgc aaatagacat aagacaagac 720
atactggaag gaaacctttc aaatgtaaaa aatgtggcaa atcattttgc atgcttttac 780
acctaatgca acataaaaaga attcatatta gagagaattc ttaccaatgt gaagaatgtg 840
gcaaagcttt taaatgggtc tcaaccctta ctagacacaa gagaattcat actggagaga 900
aacccttcaa atgtgaagaa tgtggcaaag cttttaagca gtcctcaacc cttactacac 960
ataagataat tcatactggg gagaaaccat atagatgtga agaattgtgg aaagccttca 1020
accggtcctc acaccttact acacataaga taattcatac tggagagaag ccctacaaat 1080
gtgaagaatg tggcaaagct ttttaaccagt cttcaaccct tagtacacat aagttcattc 1140
atgctggaga gaaaccctac aaatgtgagg aatgtgacaa agcttttaat cgattctcat 1200
accttactaa acataagata attcatactg gagaaaaatc ttacaaatgt gaagaatgtg 1260
gcaaaggctt taattgggtc tcaaccctta ctaaacataa aagaattcat actggagaga 1320
aaccctacaa atgtgaagtg tgtggcaaag cttttaatga gtcctcaaac cttactacac 1380
ataagatgat tcatactgga gagaaaccct acaaatgtga agaattgtgg aaagccttta 1440
accggtcccc acaacttact gcacataaga taattcatac tggagagaaa ccttacaaat 1500
gtgaagaatg tggcaaagct ttttagccagt catcaatcct tactacacat aagagaattc 1560
acactggaga gaaaccctac aaatgtgaag aatgtggcaa agcttttaac cgatcctcaa 1620
atcttactaa acataagata attcatacag gagagaaatc ttacaaatgt gaagaatgtg 1680
gtaaagcctt taaccaatcc tcaactctta ctaaacaatg gaaaattcat actagacaga 1740
aaccctacaa ctgtgaagaa tgtgacaata catttaacca gtcctcaaac cttattaaac 1800
aaaataattc atactggaga gaaactctac aaatgtcaag aatgtgggaa agcctttaag 1860
cagtcctcaa ctcttactaa gcattaaata ttggccgggt gcggtggctt atgcaaaatg 1920
gctcccagca ttttgggagg ctgaggtggg tggatcaciaa ggtcaagaga tcgagaccat 1980
cctggccaac atggtgaaac cctgtctcta ctaaaaatac aaaaattatc tgggtgtggt 2040
ggcacgtgcc tgtattccca tctactcggg aggccttaggc aggataatca cttgaacctg 2100

```

```

ggaggtggaa gttgcagtga gccaagattg taccactgca ctccagtttg gcaacagagt 2160
gagactccgt ctcaaaaaaa atttatactg tacaaaaacc ccacaagtgt gaaaaacatg 2220
gcaaagcctt ttaagaagcc ctcaattctt aacagacata agataattta tactggagag 2280
aaactctaca aatcagaaaag atgtgactac tttgacaacg cctcaaactt ttctaaccat 2340
aaaagtaatt atactggtga gaaatcctag aaatctgaag agtgagataa agccttttaa 2400
tggttgtcac acttcatgta ggtaagataa ttcatactgg ggaaaacgcc tacatgtgtg 2460
acatatggca aattatggcc ccttattgtg gaagcttttc gggataatt 2509

```

&lt;210&gt; 12

&lt;211&gt; 2119

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 70287215CB1

&lt;400&gt; 12

```

ggccaaagcc aaggtgttgg tagggctggg ctattatctg gaggtctctga gaggaatcta 60
cttatgaggt cattgaggtt gtcagtagaa ttcagctctt tgggatcagt ccagtcagtc 120
tggaacttgc atctcatatt gaaagtcctg aggattaggg caggaaatca ttagggtcac 180
cttagaattt agacaaccac agacctggtt ggtaatggga taactctgag caagtcactg 240
agaggagctt cctagagagt ccaggaagtg ccctggatcc tcaccaattt cagttttttt 300
cctgattagt caggggcttg ggggtgtcca ggatcactcc aggacatcct ttgggaacct 360
tactatggga ttcaggttgg ggtggtgtag aggattaagc tgcagaacca ctgaacatgg 420
agaaatttct gcagcagttc cccagagatg tctctgggct cctgggcagg agagaaggga 480
ctacttgcaa gtttactggg taataatgat ttcccacctt tcagggatct tctgcagaaa 540
tagcgctgga agctagagtg aggcctgagt actgccttgg cctaggatgg ctagagaatt 600
aagtgaagc acagccctgg atgccagtc tacagaagac cagatggagc ttctggtcac 660
aaaggtggag gaagaagaag ccggttttcc cagtagccca gatctgggtt ctgagggtc 720
ccgcgagcgc ttccgaggct tccgctaccc ggaggctgca gggccccgag aggcgctgag 780
tcggctccga gagctctgcc gacagtggct gcagcctgag atgcacagca aggagcagat 840
cctggagctg ctggtgctgg agcagttcct gaccatcctg ccgggggaatc tgcagagctg 900
ggtgcgggag cagcatccag agagcgggga ggaggtgggt gtgctattgg agtatttggg 960
gaggcagctg gatgagccgg cgccgcaggt agaaagaaca ggttttagtat ctgagcgctg 1020
tggcctgttt cctcctgaaa tcccagatca gattctccag gtccccgtgc ttgcccatgg 1080
aggatgctgc agagaagata aagtggtagc ttctaggctt actccagagt cccaggggtt 1140
gttgaaagtg gaagatgtgg ccctgaccct caccctgaa tggacacagc aggattcatc 1200
tcaggggaat ctctgtagag atgaaaagca ggagaacat ggcagcctgg tcttctctggg 1260
tgatgaaaaa cagactaaga gcagggactt gcctccagct gaggagcttc cagaaaagga 1320
gcatgggaag atatcgtgcc acctgagaga agacattgcc cagattccta catgtgcaga 1380
agctggtgaa caggagggca ggctacaaag aaagcagaaa aatgccacag gagggaggcg 1440
gcacatctgc catgaatgtg gaaagagttt tgctcaaagc tcaggcctga gtaaacacag 1500
gagaatccac actgggtgaga aaccctacga atgtgaagag tgtggcaaag ccttcattgg 1560
gagctctgcc cttgtcatte atcagagagt ccacactggt gagaagccat atgagtgtga 1620
agaatgtggg aaggccttca gtcatagctc agaccttate aagcatcaga gaaccacac 1680
tgggggagaag ccctatgagt gtgatgactg tgggaagacc ttcagccaga gctgcagcct 1740
ccttgaacat cacagaatcc aactgggga gaagccgtat cagtgcagta tgtgtggcaa 1800
agccttttag cgaagttcac atctcctgag acatcagagg atccatactg gggataaaaa 1860
tgttcaggaa cctgagcagg gagaggcctg gaaaagtagg atggaaagcc agttggaaaa 1920
tgttgaaact cccatgtctt ataaatgtaa tgagtgtgaa agaagtttca ctcagaatac 1980
aggcctcatt gaacatcaaa aaatccacac tgggtgagaaa ccctatcagt gtaatgcgtg 2040
tggaagaggc ttcacccgaa tttcatacct tgttcaacat cagagaagcc atgtagggaa 2100
aaacatccta tcacagtga 2119

```

&lt;210&gt; 13

&lt;211&gt; 1791

&lt;212&gt; DNA



<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 5500054CB1

<400> 13

```
gtttctccac gaggggggggt taaaggcccc caaaacatgc acacatgaaa agggcaaaaa 60
gtaaccgtcc ttgacaggga gtcttaacta gagaaggaaa cgggactaaa ctggcgggct 120
ccgtggaagc gtggccggca gcgtcccgga cgaggagcta cctgaaaact tttgttccta 180
tgcataaaga tgtctttggg ggacttggga aagaggttgc tagaagcagc aagaaaaggc 240
caagatgatg aagtgagaac gttgatggca aatggcgccc cattcaccac agactggctt 300
ggaacatcac ccctccacct tgcagctcaa tatggtcatt attccacagc agaagtactc 360
cttcgagcag gtgttagcag ggatgcccg actaaagtag acaggacccc cttgcacatg 420
gctgcagccg atggacatgc gcacatcgtg gaactgcttg ttcggaatgg tgcagatgtg 480
aatgccaaag acatgctgaa gatgacagct ttgcattggg ccacagagcg ccaccatcga 540
gatgtcgtag agttacttat caaatatgga gctgatgtcc atgctttcag caaatttgat 600
aaatcagcct ttgacatagc tctggagaaa aacaatgctg agattttggg catcctccag 660
gaagcaatgc agaatcaggt gaatgttaat ccagagagag ccaaccctgt gactgaccct 720
gtgagtatgg ctgctccatt catcttcacg tcgggtgagg ttgttaacct cgcaagcctt 780
atttcttcaa ccaacaccaa aacaacctca ggtgaccccc atgcctcaac agtacagttt 840
tcaaattcta ccacctcagt gctggctacc cttgcagctc ttgctgaggc atcagtcccc 900
ctctccaact cacacagagc cacagccaat acagaggaaa ttatagaagg aaattccgtt 960
gactcatcaa tccagcaagt aatggggagt ggaggccaga gggtcacac catagtgact 1020
gatggagtcc ctctgggtaa tatccaaact tcaatcccta ctggaggcat tggccagcca 1080
tttattgtaa ctgtgcaaga tggacagcaa gttctaactg tacttgctgg taaggttgca 1140
gaggagactg taattaaaga ggaagaagaa gagaagttgc cactaacaaa gaaaccaagg 1200
ataggagaga agacaaacag tgtggaggaa agcaaggaa gcaatgaaag agagctacta 1260
cagcaacaac tccaggaggc caatcgaaga gccaggaa accgacacca gctcctaaag 1320
aaagagcagg aagcagaaca gtaccgtctt aagctggagg ccatagcccg acagcagccc 1380
aatggagtgt atttcaccat ggttgaagag gtggctgagg tagatgctgt agtagtcaca 1440
gagggggagt tggaagagag agagacaaaa gtgactgggt cagcaggggc cacagagcct 1500
cacactagag tttccatggc aactgtttca tcttaatatg caagggccac aatttgcact 1560
gtgttcatat taatcctctt ttaaaaaagg aaatatacag aagacaaaca ttgtataaaa 1620
actaagagtg tctttaagaa gaaaactata gcaggggtaca atgcttgggc tcaggaagtt 1680
tctctgtgca actagaaaat tcaaagccat atttagggaa catTTTTTct gaggggccaa 1740
aagaataaag gaccaaattt cttagctcat atcattgctt taaacataga a 1791
```

<210> 14

<211> 1343

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 1349648CB1

<400> 14

```
ataacctgga gccggcgggc taggttggct ctttagggct tcaccccgaa gctccacctt 60
cgctcccgtc tttctggaaa caccgctttg atctcggcgg tgcgggacag gtacctcccg 120
gctgctgcgg gtgccctgga tccagtcggc tgcaccaggc gagcgagacc cttccctggg 180
ggaggctcag agttccggca ggggtgcatcc ggcctgtgtg tggcgcgagg caggggaagg 240
ggtacccggg tcttgcccc agcgtgacg ttttctctcc cctttcttct ctcttogcgg 300
ttgcggcgtc gcagacgcta gtgtgagccc ccatggcaga tacgaccccg aacggccccc 360
aaggggcggg cgctgtgcaa ttcattgatga ccaataaact ggacacggca atgtggcttt 420
ctcgcttgtt cacagtttac tgcctctgctc tgtttgttct gcctcttctt gggttgcatg 480
aagcagcaag cttttaccaa cgtgctttgc tggcaaatgc tcttaccagt gctctgaggc 540
tgcataaag attaccacac ttccagttaa gcagagcatt cctggcccag gctttgttag 600
```

```

aggacagctg ccactacctg ttgtattcac tcatctttgt aaattcctat ccagttacaa 660
tgagtatctt cccagtcttg ttattctctt tgcttcatgc tgccacatat acgaaaaagg 720
tccttgacgc aaggggctca aatagttttac ctctgctgag atctgtcttg gacaaattaa 780
gtgctaataca acaaaatatt ctgaaattca ttgcttgcaa tgaaatattc ctgatgcctg 840
cgacagtttt tatgcttttt agtgggtcaag gaagtttgct ccaacctttt atatactata 900
gattttcttac ctttcgatat tcgtctcgaa gaaacccata ttgtcggacc ttattttaatg 960
aactgaggat tgttggtgaa cacataataa tgaaacctgc ttgcccactg tttgtgagaa 1020
gactttgtct ccagagcatt gcctttataa gcagattggc accaacagtt ccatagttta 1080
acatctagtt aagctacaaa tatagtataa gcattattag cagctggtac ttctgctagg 1140
ggttgtaaat tccaggtggt acactgacct caatccaatt tacataaatt acataaattg 1200
atctcggtagg aaaaataatc attttcttgg catgttaaat caagcttaaa aagttttgag 1260
aaaatttttac tgtgctgtgt tgctaattgt taaagaagtc tgtatctagt gataaatata 1320
ccagtttttt taaaaaaaaa aaa 1343

```

<210> 15

<211> 2489

<212> DNA

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 3888314CB1

<400> 15

```

caggcgcact tccagttctg cccaccgct gcggccattg tccgaccccg gtgcggctga 60
ggcccctttg ggcagccct cgcagatca gaattggaga caaccgagcc ttcggcgggg 120
gcgggagagg gtcgccgagg ctctgggtgg gccggaggtc gcgaaatccg gagcccccca 180
gaggcggtag ttctgactcc tgccccggag ccggggcctg gcgaggcagg aatggccccg 240
aggcctccga ccgcgcgcgc ccaggaatca gtgacattca aagatgtgtc tgtggacttc 300
accaggaag aatggtacca tgtcgacct gctcagagga gcttatacag ggatgtgatg 360
ctggagaact atagccacct ggtttctctt ggatatcaag tttccaagcc agaggtgatc 420
ttcaaattgg agcaaggaga agagccatgg atatcagagg gagaaatcca acgacctttc 480
tatccagact ggaagaccag gcctgaagtc aaatcatcac atttgcagca ggatgtatca 540
gaagtatccc actgcacaca tgatctctta catgctacat tagaagactc ctgggatgtt 600
agcagccagt tagacaggca acaggaaaac tggaagagac atctgggatc agaggcatcc 660
accagaaga aaataattac accacaagaa aattttgagc aaaataaatt tggtgaaaat 720
tctagattga acaccaatth ggttacacaa ctgaacattc ctgcaagaat aaggcctagt 780
gaatgtgaga cccttggaag caatttgga cataatgcag acttacttaa tgagaataat 840
attcttgcaa aaaagaaacc ctataagtgt gataaatgta gaaaagcctt tattcataga 900
tcatcgctta ctaaaccatga gaaaacacat aaaggagagg gagctttccc taatggaaca 960
gatcaaggaa ttatccttgg aaagaaacac catgaatgta ccgactgttg gaaaaccttt 1020
ctctggaaga cacagcttac tgagcatcag agaattcaca ctggggagaa accctttgaa 1080
tgcaatgtat ttggaaaagg cttcaggcat agctcatctc ttggtcagca tgagaatgct 1140
cataccggag agaaacccta tcagtgtagt ctctgtggga aagccttoca gcgcagctcc 1200
tccttggttc aacaccagcg aattcacact ggagagaaac cctatcgatg taatctatgt 1260
gggagggtcct ttaggcattg cacatccctc actcaacacg aggtcacaca cagtggagag 1320
aagcccttcc agtgtaagga atgtgggaaa gccttttagtc gatgttcttc ccttgtccaa 1380
catgagagga ctcatactgg agagaaacct tttgaatgta gcataatgtg gagggctttt 1440
ggtcagagcc catcccttta taaacatatg aggattcata agagaggcaa accttaccaa 1500
agcagtaact acagcataga tttcaagcac agcacatctc tcaactcagga tgaaagcact 1560
cttaccgaag tgaaatccta ccattgtaat gactgtgggg aagactttag tcacattaca 1620
gactttactg accatcagag gatccatact gcagagaacc cctatgattg tgagcaggct 1680
tttagtcagc aagctatthc tcatcctgga gagaaacct atcaatgtaa tgtatgtggg 1740
aaagctttca aaaggagtac aagtttcata gagcatcaca gaattcatac tggagagaaa 1800
ccctatgaat gtaatgagtg tggagaagcc tttagtcgac gctcatcgct tactcaacat 1860
gagagaacct acactggaga gaaacctat gaatgtattg actgtgggaa agcctttagt 1920
caaagttcat ctctcattca gcatgagaga actcatactg gagagaagcc ctatgaatgt 1980
aatgaatgtg ggagagcctt ccgaaaaaaaa accaacctgc atgatcatca gagaattcat 2040

```

actggagaaa	aaccctattc	ttgtaaggaa	tgtgggaaaa	acttcagccg	aagttcagct	2100
cttactaaac	accagagaat	tcatactcga	aataaactct	aggaaccgtg	aaattaagga	2160
atttgagaaa	tgcttttagct	aaaatgttct	gattcaggat	cagaggattc	ttagagagct	2220
tgggaatgta	atgaattacg	tgtgtgttta	tacgttgtgt	gtggagaaaa	ctgccagtag	2280
acagattttt	tttttttttt	taacataaag	acacattctc	agatctgatt	acagactagt	2340
gtaaaaacag	ctacatgtat	gtagctgggt	ggggatgata	tgcctgtatg	ttggactttg	2400
cttttgaata	tatgtatgca	ggatatcatc	aagtttcaac	atcttgactt	gtgaccccca	2460
atgtcaacag	ctttttttaa	aaacaaatt				2489

&lt;210&gt; 16

&lt;211&gt; 6236

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 3276670CB1

&lt;400&gt; 16

gagacagaca	gacagccaga	gagagcggga	gggaggaggg	cggaggagga	ggagcaggag	60
gactacgcgg	ggcgcggtcc	ctcccagcag	ccgggcgccc	gcgtcgctgc	caccgccgga	120
agacagacgt	ccccaaaaac	ccctcgctcg	gctcggaccc	acggagccgg	cagcccagca	180
ctccgatggg	agcttgctcg	attgtcaacc	catcggcctg	gaccccagag	cccaccagct	240
tgagcgtgtg	cagagggccc	cgagtgcggg	gcgacgctca	ggcgttgca	gagaccaccc	300
tggagatgct	gatgaagaca	ggcctttcct	gctgagccca	gcctatcttc	gcagcctgct	360
cagaaaaggg	aactgaggct	catggatata	cagtgaagtc	tacctcctag	cctgaggact	420
cttccctgcg	ccttgatgac	ctgagagaa	tgggaggaga	aaggggttgt	gacaatgacc	480
aggactcagg	ggcagagggt	gaaaccagag	tgcctgatgt	gtgcagcttg	ctgaccccc	540
caggacacct	ctctaccgga	gtctgtggaa	gcagagaact	gggaggccaa	ggcgctgtcc	600
acgccccgag	gcctgggctc	aggccgaagt	tgctccacag	cacccccagc	tgacagccag	660
cgactggcag	agcccccagg	cacggggggc	ctctctcttc	cagagatctc	cagtgtgaca	720
gcccaggcgg	cagagccgag	ggtcctgggt	ccggcttctc	gcacctcat	gtcagccccg	780
gccccgccc	gcggcctgcg	gaggacaagg	tcaggatgcg	tgtgaagccc	cagggcctgg	840
tggtgacttc	cagtgcctgt	tgcagctctc	ctgactacct	ccgggaaccc	aagtactacc	900
ccggcgggcc	ccccaccccc	cggccccctg	ttcccacccg	gccccctgcc	agcccacctg	960
acaaggcctt	ctccacccac	gccttctccg	agaacccacg	cccacccccca	cgccggggacc	1020
ccagcacccg	gcgcccacca	gtccttgcca	agggggacga	cccgtgccc	ccacggggcag	1080
cccgtcctgt	ctcacaggcc	cgctgccccca	caccggctcg	agacggcagc	agctcccgcac	1140
gctgctggga	caacgggcgg	gtgaacctgc	gaccagtggg	gcagctgatt	gacatcatga	1200
aggacctgac	acgcctctcc	caggacctgc	agcacagtgg	tgtacacctg	gactgtgggt	1260
ggctccgact	gagccgccc	cctgcaccgc	caccggcgca	cctacagtac	agcttctctc	1320
cctcaccag	tttggccaac	agcatccgta	gccctgagga	gcggggccacc	ccacacgcca	1380
agtcggagcg	gcccagccat	cccctctacg	agcctgagcc	tgagcctagg	gacagtcccc	1440
agcctggcca	aggccatagt	cccggagcca	cggctgcggc	cacgggtctg	ccccagagc	1500
ctgagccaga	cagcactgat	tactcagaac	ttgctgacgc	cgacatcctt	agtgaactgg	1560
cctccctcac	ttgcccagag	gcccagctgc	tagaggccca	ggccctcgag	ccaccatcgc	1620
ccgagccgga	gcctcagctc	ctggaccccc	agccccgctt	cctggacccg	caggcactag	1680
agccgctcgg	ggaagctctg	gagctgccac	ccctgcaacc	tcttgctgat	cctctggggc	1740
tgccggggcct	ggctctccag	gccctggaca	ccctgcctga	ctccttgagg	tcgcagctgc	1800
ttgaccccca	ggcactcgac	cccctgcccc	agctgcttga	cgtcccaggt	cgccgtctgg	1860
agccccagca	gcccctgggg	cactgcccac	tggccgagcc	cttgccgctg	gacttgtgtg	1920
caccgcacgg	cccccccggg	cctgagggtc	accccaagta	cgccttgccg	cgcaactgata	1980
ggccaaaagat	cctgtgtcgc	cggcggaag	ccggacgggg	acgcaaggca	gacgccggac	2040
ccgaggggcc	cctactgccc	ctgcctatgc	ccacgggggt	ggtggctgcc	ctggccgaac	2100
ccccaccacc	accgctctct	ccacccccct	ccctgccagg	cccaggcccc	gtctcagctc	2160
cagagttgaa	gccggaatct	tcccagaccc	cagtgggtct	taccgcaaaa	ggcaagtggc	2220
ggggcggtgc	gcgcatgggt	gtgaagatgg	ccaagatccc	cgtatcgctg	gggcccggga	2280
acaagaccac	atacaaagtg	tcttccttga	gcagcagcct	gagcgtggag	ggcaaggagc	2340

tgggcctgcg	cgtgtcggct	gagcccaccc	cgctgctgaa	gatgaagaac	aatgggcgga	2400
acgtggtagt	ggtcttccca	cccggtgaga	tgcccattat	tctcaaactg	aagcgcgcc	2460
gccctcctaa	gaacctgctg	ctgggtcccc	gcaagcccaa	ggagccagct	gtgggtggcg	2520
ccgaggcagc	cactgtggca	gcgggcacca	tggccatgcc	agaggtaaaa	aaacgacggc	2580
ggcggaagca	gaagctggca	tctccccagc	catcctatgc	agcagacgcc	aacgacagca	2640
aggccgagta	ctcagacgtc	ctggccaagc	tggccttcc	gaaccgccag	agccagtgcg	2700
ctggacgggtg	ctcaccgccc	cgctgctgga	caccagtgta	gccggagtcg	gtgcaccagg	2760
ccccgcacac	ccagagcatc	tcccacttcc	tgcctcgtgt	gcagggttc	cgcgggcgtg	2820
ggggcaaagc	aggcgtttt	ggtggccggg	gtgggggcca	tgcggccaag	tcagcccgat	2880
gctccttcag	tgacttctt	gagggcatcg	gcaagaaaa	gaagtggtg	gccgtggcag	2940
ccgctggggg	cggggggccc	ggccttactg	agtgggggca	cccacgcaaa	cggggccggg	3000
gggaggtaga	cgctgtgact	gggaagccaa	agcgcaagag	acgggtcccg	aagaatggga	3060
ctctgttccc	agagcagggtg	cccagtggcc	caggctttgg	ggaggcagg	gctgagtggg	3120
ccgggggataa	gggtggtggc	tggggccctc	accatgggca	cccaggcgga	caagctggcc	3180
gaaactgtgg	gtttcagggg	accgaggccc	gggcctttgc	ctccactggg	ctggagagt	3240
gagcctcagg	ccgtggcagc	tactacagca	cggtgtgcac	ctcaggccag	accgagctca	3300
gccaggagcg	ccaaaacctc	ttcaccggct	actttcgctc	gctgctcgat	tcggatgact	3360
cctccgatct	cttggacttt	gccctctcag	cctctcgccc	agagtcccgg	aaggcatcgg	3420
gcacctatgc	agggccaccc	accagtggcc	tgcttgccca	gcggggcctg	gccaccttcc	3480
ctagccgggg	agccaaggcc	agcccagtgg	cagtgggtag	cagcggggct	ggggcgggacc	3540
cctcctttca	gcctgtcctg	tccgcgcgcc	agaccttccc	accaggacga	gcagcaagct	3600
atgggctaac	tccagccact	tcagactgcc	gggcagccga	gaccttcccc	aagctggtgc	3660
ccccgccctc	agccatggcc	cgctcaccta	ccaccaccc	gcctgccaac	acctacctgc	3720
cccagtacgg	cggctatggg	gccggacaaa	gcgtattcgc	cccaactaag	ccctttacag	3780
gccaggactg	cgctaacagc	aaggactgca	gcttcgccta	tggcagtggc	aacagcctcc	3840
ctgcctcacc	cagcagcgcc	cacagcgccg	gctatgcccc	accgcctacc	ggggggccct	3900
gctgcctacc	aagcaaggcc	tccttcttca	gcagctctga	ggggggcccc	ttctctggtt	3960
cagccccac	gccccctgcg	tgtgacagcc	gggcagcac	agtctcgccc	ggtggctaca	4020
tggtacccaa	gggcaccaca	gcctctgcca	cctctgcagc	ctctgcgcgc	tcctcctcct	4080
cctcctcctt	ccagccctcg	cccagagaact	gtcggcagtt	tgcgggggct	tctcagtggc	4140
ctttccggca	gggctatgga	ggcctggact	gggcctcaga	ggcctttagt	cagctctaca	4200
atcccagttt	tgactgccac	gtcagcgagc	ccaacgtgat	cctggacatc	tccaactaca	4260
caccgcagaa	ggtgaagcag	cagacggctg	tgtcggagac	cttctctgag	tcctcctccg	4320
acagcaccca	gttcaatcag	ccggttggtg	gcgggggggt	tcggcggtgc	aacagcgagg	4380
cctcaagtag	tgagggccag	tcgagcctgt	ccagcctgga	gaaactgatg	atggactgga	4440
acgaggcatc	atctgcccc	ggctacaact	ggaaccagag	tgtcctcttt	cagagtagct	4500
ccaagccggg	ccgtggacgg	cggaagaagg	tggacctgtt	cgaggcctca	catctgggct	4560
tcccagacatc	cgcctctgcc	gctgcctcag	gctacccatc	caaacggagc	actggggccc	4620
ggcagccgcg	aggtggacgg	ggcggtgggg	cctgctcagc	caagaaggag	cggggtggcg	4680
cagcgcccaa	agccaagttc	atccccaaag	cacagccagt	caaccactg	ttccaggaca	4740
gtcctgacct	cggcctggac	tactatagcg	gggacagcag	catgtcacca	ctgccctcac	4800
agtcgagggc	cttcggcggtg	ggagagcgag	accctgtgta	cttcatagga	ccctactcca	4860
tgaacccgtc	cacgccttcc	gatggcacct	ttggccaagg	cttccactgc	gactcgccca	4920
gctgggtgc	tcccagactt	gatggcaagc	atttccacc	gctggggccc	ccaccacggg	4980
tgtttgacgc	cggcctgcag	aaggcatact	cgcccactg	ctcgccctaca	ctgggcttca	5040
aggaagagct	gcgggccaccg	cccacaaaag	tggctgcctg	cgagccctc	aagcatggac	5100
tccagggggc	cagcctgggc	cacgcagctg	cagcccaggc	ccacctgagc	tgccggggacc	5160
tgccgctggg	ccagccccac	tacgattccc	ccagctgcaa	gggcacagcc	tattgggtacc	5220
ctccaggctc	agctgcccgc	agccccccct	atgaaggcaa	ggtgggtaca	gggctgctgg	5280
ctgacttcc	gggcaggacg	gaggccgcgt	gcctcagtgc	ccctcacctg	gctagccac	5340
cagccacgcc	caaggccgac	aaggagccac	tggaaatggc	ccggccccct	ggcccccccc	5400
gtggccctgc	tgcagccgct	gctggctatg	gctgcccact	ccttagtgac	ttgacctgt	5460
cccccgctgc	gagggactcg	ctgctgcccc	tgcaggacac	cgcctacagg	taccagggt	5520
ttatgcccc	ggcgcatcct	ggcctgggtg	ggggccccaa	gagcggcttc	ctggggccca	5580
tggcggaacc	tcaccccgag	gacacattca	ccgtcacatc	cctgtagtgc	caactgaagt	5640
gccgactgga	ccgcgaggtt	ttgttcctgg	ctttcagaaa	accaacgcca	agatccctcc	5700
cagcgtccac	atcgctcctc	ggcaggagct	cctgccccct	tgctccccc	cctgccccct	5760
acacccccctg	cagacccatc	tccctccacc	ccctccccc	catctcctcc	acgcagaagc	5820

cgaagggtgag	cccttttctgc	acaaaaccag	caattgtaaa	tacttttttaa	aaatgtacaa	5880
aacttaaaaa	caaaacacag	ttttagaaaa	agacaaaaaa	aaaaaagaga	gagagagagc	5940
gagagagcga	gcgtgtgcaa	gagggtgcga	gcggggcccc	gagggtgtcca	gagcccctgc	6000
aagtatgcac	tgagaaattt	atctacaggt	cgtttgacaa	aaatgaacaa	tatcctattt	6060
attgtatata	ctgtttttatt	ataaatcgtg	gattgtatat	tgcattctgt	aaacctgctg	6120
tggtccgtgg	tgtgcaattc	gcatgtctag	tgggatggag	aagaactctt	gcggacgttt	6180
gtttttcttc	ctatttttcc	tcttttgggt	ttccctcctg	tggtgtgtgt	ggttttt	6236